Conjugacy Rigidity, Cohomological Triviality, And Barycentres of Invariant Measures

by

Oliver Mark Jenkinson

A thesis submitted to the University of Warwick
for the degree of Doctor of Philosophy

Mathematics Institute, University of Warwick

August 1996

Table of Contents.

Acknowledgments

First and foremost I would like to thank my supervisor Mark Pollicott for all his help and encouragement, for suggesting interesting problems to work on, and for numerous enlightening conversations. It was a real pleasure and privilege to learn so much interesting mathematics from Mark.

I am also grateful to Hans Henrik Rugh for taking such an interest in my work, and for critically reading some of the material in this thesis.

I would like to thank all my friends and colleagues at the Mathematics Institute, for providing such a convivial working environment. In particular I thank Charles Walkden for many interesting discussions on dynamics, and Paul Sanders for his advice on group theory and all things computer-related.

This thesis has also benefited greatly from conversations and correspondence with Jim Anderson, Viviane Baladi, Giovanni Cutolo, André Rocha, and Richard Sharp.

Finally, I would like to thank the Engineering and Physical Sciences Research Council for funding this research.

Declaration

The material in this thesis is, to the best of my knowledge, original except where stated otherwise. Particular attention should be drawn to sections 1·7 and 1·8, which are entirely expository.

Summary

This thesis consists of three chapters and four appendices. Each chapter is a self-contained study of one topic in dynamical systems. Each chapter has its own notation.

In chapter 1 we study conjugacies h between piecewise smooth conformal expanding Markov maps T_1, T_2. We prove that if h is piecewise continuous, satisfies an absolute continuity condition, and has essentially bounded partial derivatives, then it is as smooth as T_1, T_2. We use this result to give a new proof of part of Mostow's theorem on the rigidity of manifolds of constant negative curvature.

In chapter 2 we study the cohomology of certain two-dimensional subshifts X with a semi-safe symbol. The simplest examples of such subshifts are the full shift and the golden mean shift. We prove the triviality of all locally constant cocycles on X taking values in a locally (residually finite) group. For real-valued cocycles we extend this result to the Hölder category.

In chapter 3 we consider the doubling map of the circle, and study the convex set Ω of barycentres of invariant measures. We prove that each interior point of Ω is the barycentre of an equilibrium state of a particular kind, and that this equilibrium state maximises entropy over all measures with this barycentre. We prove that any measure whose barycentre lies on the boundary $\partial\Omega$ is not fully supported, and conjecture that its support has zero Hausdorff dimension. We conjecture further that $\partial\Omega$ is non-differentiable at a countable dense set of points, the worst possible regularity for the boundary of a planar convex figure.

Appendix A contains the proof of a technical lemma stated in section 1·3. Appendix B contains an alternative proof (due to de la Llave, Marco & Moriyón) of a theorem stated and proved in section 1·5. Appendix C contains numerical data to support the conjectures of chapter 3. Appendix D is a graphical plot of the data contained in Appendix C.

Chapter 1. Conjugacy Rigidity For Piecewise Smooth Conformal Expanding Markov Maps

Section 1·1. Introduction.

Shub & Sullivan [66] proved that if two C^r, $r \geq 2$, expanding maps of the circle are absolutely continuously conjugate, then in fact they are C^r conjugate.

In the context of hyperbolic dynamics this kind of *rigidity* result, where a weak assumption about the conjugacy implies a strong one, has been the motif of several recent articles. For example, the regularity of conjugacies between low dimensional C^∞ Anosov systems was investigated in a series of papers by de la Llave, Marco & Moriyón [33], [35], [37], [38]. Amongst other things, they proved that the eigenvalues of the derivatives at periodic points form a complete set of invariants for C^∞ conjugacy of C^∞ Anosov diffeomorphisms of the two-dimensional torus. A consequence of this is that any Lipschitz conjugacy between two C^∞ Anosov diffeomorphisms of the two-dimensional torus is actually a C^∞ conjugacy. In a similar vein, Arteaga [2] (see also [3] for circle maps) proved that a topological conjugacy h between two C^r, $r \geq 1$, expanding maps T_1, T_2 of a compact manifold X is absolutely continuous if and only if

$$|\det(DT_1^n)(x)| = |\det(DT_2^n)(h(x))|$$

for all $x \in X$ with $T_1^n x = x$, $n \geq 1$. Combining this result with the theorem of Shub & Sullivan mentioned above, we obtain a complete set of invariants for C^r conjugacy between C^r expanding circle maps, in the case $r \geq 2$.

In this chapter we present a generalisation of Shub & Sullivan's result to higher dimensions. We consider conjugacies h between piecewise smooth conformal expanding Markov maps $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$. The spaces X_i are partitioned manifolds (the terminology is explained in §1·2), and the smoothness assumption on the T_i is that the derivative should be (piecewise) Hölder. The simplest examples of such maps are precisely the smooth expanding circle maps. We prove (Theorem 1·37) that if the conjugacy is (piecewise) continuous, differentiable almost everywhere with essentially bounded partial derivatives, and satisfies an absolute continuity condition, then in fact it is as (piecewise) smooth as the maps themselves.

The principal class of piecewise smooth conformal expanding Markov maps arise in the context of Kleinian group actions on S^n, and are known as Bowen-Series maps. These were introduced by Bowen & Series [8] in the case $n = 1$, and we sketch their construction in §1·7. In §1·8 we indicate how this construction generalises to higher dimensions. The case $n = 2$ has recently been studied by André Rocha [59].

An application of our conjugacy rigidity result (Theorem 1·37) to Bowen-Series maps leads to a new proof of part of Mostow's Rigidity Theorem, which we present in §1·9. This well-known theorem of geometry states that for $n \geq 3$, if two hyperbolic n-manifolds M_1 and M_2 have the same homotopy type, then in fact they are isometric. As in Mostow's original proof (see Mostow [42]) we lift the homotopy equivalence to a homeomorphism $\tilde{H}$ of S^{n-1} (the boundary of the universal cover of both M_1 and M_2). We then observe that $\tilde{H}$ conjugates two Bowen-Series maps, and also satisfies the hypotheses of Theorem 1·37. We deduce that $\tilde{H}$ is smooth, and from this that it is conformal. Then in the usual way we extend $\tilde{H}$ to an isometry on the universal cover, and push it down to an isometry between M_1 and M_2.

The proof of Theorem 1·37 uses ideas contained in a recent paper by Bill Parry and Mark Pollicott [49] concerning the regularity of solutions to the unitary-valued measurable cocycle equation. This kind of problem was first addressed by A. Livsic [31], [32], in the context of Anosov diffeomorphisms and subshifts of finite type. In contrast to the techniques of Livsic, Parry & Pollicott exploit the spectral properties of certain bounded linear operators, known as Ruelle-Perron-Frobenius operators, which act on Hölder functions defined on a subshift of finite type. In our context the Ruelle-Perron-Frobenius operators are defined in terms of a piecewise smooth expanding Markov map $T : X \to X$ of a partitioned manifold, and act on the space of piecewise smooth vector-valued functions. The quasi-compactness (Theorem 1·27) of these operators allows us to prove (Theorem 1·30) that any L^1 solution to the smooth unitary-valued cocycle equation has a smooth version. This kind of quasi-compactness result was first proved by Pollicott [52], and in the smooth case is due to Tangerman [68].

In §1·2 we present some background material on Ruelle-Perron-Frobenius operators, together with a review of some spectral theory. In §1·3 we introduce an operator defined in terms of a smooth, unitary matrix-valued function. We call this the *matrix operator*, and investigate its spectral properties. The key results are that its iterates are uniformly bounded in the L^1 norm, and that its restriction to C^k functions is quasi-compact.

In §1·6 we make the additional assumption that our map T is conformal, and suppose that it is conjugate to some other piecewise smooth conformal expanding Markov map, where the conjugacy h is differentiable almost everywhere. Differentiating the conjugacy equation leads to a cocycle equation, where the derivative Dh is an almost everywhere solution. Our aim is to show that in fact Dh is a solution *everywhere*, and is optimally smooth. The conformality assumption allows us to split the cocycle equation into two separate equations, one matrix-valued, the other real-valued. The results of §1·4 are brought to bear on the matrix-valued equation, while we use the more classical techniques (which we summarise in §1·5) to deal with the real-valued equation. The trick in the proof of Theorem 1·37 is to use a 'bootstrap of regularity' argument, where at each step we deduce one further degree of differentiability of the conjugacy.

Section 1·2. Definitions and Preliminary Results.

In this section we fix our notation and define certain spaces of functions. We collect some elementary results from measure theory, and some standard facts about unitary matrices. We also review the spectral theory of bounded linear operators, in particular the theory of Ruelle-Perron-Frobenius operators.

We will be considering several different normed vector spaces, so we will carefully label the various norms to avoid confusion.

The vector space $\mathbf{C}$ will be given its usual absolute value $|\cdot|$.

Let us fix the integer $d \geq 1$. We will consider certain d-dimensional manifolds, and also certain groups of $d \times d$ matrices. Many of our results do not require that the dimension of the manifold is the same as the size of the matrices, though in §1·6 we look at derivatives (considered as matrices) of maps between manifolds, and here they *are* the same.

$|\cdot|_d$ will denote the Euclidean norm on $\mathbf{C}^d$.

Let $M(d)$ denote the space of $d \times d$ matrices with entries in $\mathbf{C}$. As a vector space, $M(d)$ is isomorphic to $\mathbf{C}^{d^2}$. Let $|\cdot|_{d^2}$ be the operator norm on $M(d)$ given by

$$|A|_{d^2} = \sup\{|Av|_d : v \in \mathbf{C}^d \text{ with } |v|_d = 1\}.$$

Then $|\cdot|_{d^2}$ can be considered as a norm on $\mathbf{C}^{d^2}$.

Similarly, by identifying $\mathbf{C}^{d^4}$ with the space $M(d^2)$, we can define $|\cdot|_{d^4}$ to be the norm induced by the operator norm on $M(d^2)$. That is,

$$|A|_{d^4} = \sup\{|Av|_{d^2} : v \in \mathbf{C}^{d^2} \text{ with } |v|_{d^2} = 1\}.$$

Let $< \cdot, \cdot >_d$ denote the Euclidean inner product on $\mathbf{C}^d$.

Let $< \cdot, \cdot >_{d^2}$ denote the Euclidean inner product on $\mathbf{C}^{d^2}$.

Then $< \cdot, \cdot >_{d^2}$ is an inner product on $M(d)$, and we have the following standard results.

Lemma 1·1. *The inner product on $M(d)$ is given by the formula*

$$< A, B >_{d^2} = \text{Trace}(AB^\star),$$

where $B^\star$ is the adjoint (i.e. conjugate transpose) of the matrix B. $\square$

Lemma 1·2. *$\text{Trace}(CD) = \text{Trace}(DC)$ for all $C, D \in M(d)$.* $\square$

Let $U(d)$ denote the space of $d \times d$ **unitary** matrices. That is,

$$U(d) = \{A \in M(d) : A^\star A = AA^\star = I\}.$$

We equip $U(d)$ with the norm $|\cdot|_{d^2}$ and inner product $< \cdot, \cdot >_{d^2}$.

Similarly, let $U(d^2)$ denote the space of $d^2 \times d^2$ unitary matrices. That is,

$$U(d^2) = \{A \in M(d^2) : A^\star A = AA^\star = I\}.$$

We equip $U(d^2)$ with the norm $|\cdot|_{d^4}$ and inner product $< \cdot, \cdot >_{d^4}$.

Note that both $U(d)$ and $U(d^2)$ are compact groups under matrix multiplication.

It will be useful to note a few elementary properties of unitary matrices.

Lemma 1·3.

(i) Let $A \in M(d)$. Then

$$A \in U(d) \iff < Ax, Ay >_d = < x, y >_d \text{ for all } x, y \in \mathbb{C}^d.$$

(ii)Let $A \in M(d^2)$. Then

$$A \in U(d^2) \iff < Ax, Ay >_{d^2} = < x, y >_{d^2} \text{ for all } x, y \in \mathbb{C}^{d^2}. \quad \square$$

As a corollary we obtain that every unitary matrix has unit norm.

Lemma 1·4.

(i) $A \in U(d) \Rightarrow |A|_{d^2} = 1$,

(ii) $A \in U(d^2) \Rightarrow |A|_{d^4} = 1$. $\square$

Definition 1·1. *A d-dimensional* **partitioned manifold** X *is a disjoint union of a finite collection* $\mathcal{P} = \{P_1, \ldots, P_n\}$ *of d-dimensional compact connected* C^∞ *manifolds with piecewise* C^∞ *boundaries. We assume that each* P_i *lies in some ambient Euclidean space* $\mathbf{R}^N$, *so that the metric on* X *is Euclidean. We let* $\partial\mathcal{P} = \partial P_1 \cup \ldots \cup \partial P_n$ *denote the union of the boundaries. We assume that if* $P_i \neq P_j$ *then* $int(P_i) \cap int(P_j) = \emptyset$. *We identify each tangent space* $\tau_x X$ *with Euclidean space* $\mathbf{R}^d$.

In particular, X *is a metric space with distance function* ρ *induced from the Euclidean norm on the tangent spaces.*

If Y is a subset of some finite dimensional normed vector space Z, then let $F(X, Y)$ denote the space of all functions $X \to Y$.

We will be particularly interested in the case where

$$Y = \mathbf{R}^+,\ \mathbf{R},\ \mathbf{C},\ \mathbf{C}^d,\ \mathbf{C}^{d^2},\ \mathbf{C}^{d^4},\ U(d), \text{ or } U(d^2).$$

We will consider differentiable functions w defined on X. The derivative at the point x is denoted $D_x w$, while the derivative map $x \mapsto D_x w$ is denoted Dw. The j^{th} derivative map, which we think of as a multilinear map, is denoted $D^j w$.

Let $C^{\mathbf{k}}(X, Y)$ denote the space of piecewise $C^{\mathbf{k}}$ functions $w : X \to Y$. By a piecewise $C^{\mathbf{k}}$ function we mean one that is $C^{\mathbf{k}}$ on the interior of each piece P_i of the partition $\mathcal{P}$, and whose appropriate directional derivatives exist on each boundary $\partial\mathcal{P}_i$. Since we are thinking of $\mathcal{P}$ as a disjoint union, we do *not* require that $w|_{\partial\mathcal{P}_i} = w|_{\partial\mathcal{P}_j}$ when $\partial\mathcal{P}_i \cap \partial\mathcal{P}_j \neq \emptyset$.

Here the regularity index $\mathbf{k}$ is an integer $k \geq 0$ (in which case we also write $\mathbf{k} = (k, 0)$), or $\mathbf{k} = (k, \epsilon)$ with integer $k \geq 0$ and $0 < \epsilon \leq 1$ (we say that a function is $C^{(k,\epsilon)}$ if it is C^k and if the k^{th} derivative is Hölder continuous of exponent ϵ), or $\mathbf{k} = \infty$ (in which case we also write $\mathbf{k} = (\infty, 0)$). We write $|\mathbf{k}| = k$ in the first case, $|\mathbf{k}| = k + \epsilon$ in the second case, and $|\mathbf{k}| = \infty$ in the third case.

We can define a lexicographic order on the regularity indices as follows:

Let $\mathbf{k} = (k, \epsilon)$ and $\mathbf{k}' = (k', \epsilon')$.

We write $\mathbf{k} < \mathbf{k}'$ if either (i) $k < k'$, or (ii) $k = k'$ and $\epsilon < \epsilon'$.

We write $\mathbf{k} \leq \mathbf{k}'$ if either $\mathbf{k} < \mathbf{k}'$ or $\mathbf{k} = \mathbf{k}'$.

With this order it is clear that:

$$\mathbf{k} < \mathbf{k}' \quad \Rightarrow \quad C^{\mathbf{k}'}(X, Y) \subsetneq C^{\mathbf{k}}(X, Y).$$

The derivative of a piecewise $C^{\mathbf{k}}$ function, $\mathbf{k} \geq \mathbf{1}$, is a piecewise $C^{\mathbf{k}-1}$ function. Here

$$\mathbf{k} - 1 = \begin{cases} (k-1, \epsilon) & \text{if } \mathbf{k} = (k, \epsilon) \text{ is finite;} \\ \infty & \text{if } \mathbf{k} = \infty. \end{cases}$$

Let $\| \cdot \|_\infty$ denote the supremum norm on the space $C^{\mathbf{k}}(X, Y)$.

That is, if $w \in C^{\mathbf{k}}(X, Y)$ then we define

$$\| w \|_\infty = \sup\{|w(x)|_Y : x \in X\}$$

where $| \cdot |_Y$ is the norm on Y (induced from the norm on the vector space Z).

Let $\| \cdot \|_\epsilon$ denote the ϵ-Hölder norm on the space $C^{(0,\epsilon)}(X, Y)$.

That is, if $w \in C^{(0,\epsilon)}(X, Y)$, then we define

$$\|w\|_\epsilon = \sup \left\{ \frac{|w(x) - w(y)|_Y}{\rho(x, y)^\epsilon} : x, y \in P_i \text{ for some } 1 \leq i \leq n,\ x \neq y \right\}.$$

If $k < \infty$ then let $\| \cdot \|_k$ denote the C^k norm on the space $C^{\mathbf{k}}(X, Y)$. That is, if $w \in C^{\mathbf{k}}(X, Y)$, then we define

$$\| w \|_k = \sum_{j=0}^{k} \| D^j w \|_\infty .$$

If $\mathbf{k} = (k, \epsilon) < \infty$ then let $\| \cdot \|_{\mathbf{k}}$ denote the $C^{\mathbf{k}}$ norm on the space $C^{\mathbf{k}}(X, Y)$. That is, if $w \in C^{\mathbf{k}}(X, Y)$, then we define

$$\|w\|_{\mathbf{k}} = \|w\|_k + \|D^k w\|_\epsilon.$$

For $\mathbf{k} < \infty$ the space $C^{\mathbf{k}}(X, Y)$ is a Banach space with respect to the $C^{\mathbf{k}}$ norm $\| \cdot \|_{\mathbf{k}}$. The space $C^\infty(X, Y)$ does not have a natural Banach norm. Henceforth, whenever we write $C^{\mathbf{k}}(X,Y)$ it will be understood that $\mathbf{k} < \infty$, and that the space is equipped with its Banach norm.

We now review the spectral theory of bounded linear operators acting on the Banach space $C^{\mathbf{k}}(X,Y)$ (in fact these standard results hold for *any* Banach space). A good reference is Reed & Simon [55].

If $M : C^{\mathbf{k}}(X,Y) \to C^{\mathbf{k}}(X,Y)$ is a bounded linear operator, then we define the operator norm $\|\!| \cdot \|\!|_{\mathbf{k}}$ by

$$\|\!| M \|\!|_{\mathbf{k}} = \sup\{\| Mw \|_{\mathbf{k}} : w \in C^{\mathbf{k}}(X,Y) \text{ with } \| w \|_{\mathbf{k}} = 1\}.$$

Recall that the *spectrum* $\sigma(M)$ of a bounded linear operator $M : C^{\mathbf{k}}(X,Y) \to C^{\mathbf{k}}(X,Y)$ is defined as:

$$\sigma(M) = \{\lambda \in \mathbf{C} : \lambda I - M \text{ is not a bijective operator}\}.$$

We define $r(M) = \sup_{\lambda \in \sigma(M)} |\lambda|$ to be the *spectral radius* of M. We will need the following result.

Lemma 1·5. *Spectral radius formula*

If $M : C^{\mathbf{k}}(X,Y) \to C^{\mathbf{k}}(X,Y)$ is a bounded linear operator, then its spectral radius is given by

$$r(M) = \lim_{n \to \infty} \|\!| M^n \|\!|_{\mathbf{k}}^{1/n}.$$

Proof. See Reed & Simon [55], page 192. □

Any $\lambda \in \mathbf{C}$ satisfying $Mw = \lambda w$ for some $w \in C^{\mathbf{k}}(X,Y) \setminus \{0\}$ is clearly an element of $\sigma(M)$. Such a λ is called an *eigenvalue* of M, and the function w is a corresponding *eigenfunction*. The *multiplicity* of λ is the dimension of the generalised eigenspace $\cup_{n \geq 1} \text{Ker } (\lambda I - M)^n$.

We obtain the *essential spectrum* $\sigma_{ess}(M)$ by removing from $\sigma(M)$ all isolated eigenvalues of finite multiplicity.*

* A more general definition of the essential spectrum considers the operator M modulo compact operators (which form a maximal ideal in the set of bounded operators, so that the quotient is well-defined). However, this approach is unnecessary in the present context.

We define $r_{ess}(M) = \sup_{\lambda \in \sigma_{ess}(M)} |\lambda|$ to be the *essential spectral radius* of M. We will need the following result.

Lemma 1·6. *(Nussbaum, [44]) Essential spectral radius formula*

If $M : C^{\mathbf{k}}(X,Y) \to C^{\mathbf{k}}(X,Y)$ is a bounded linear operator, then its essential spectral radius is given by

$$r_{ess}(M) = \lim_{n\to\infty} [\inf\{ \| M^n - K \|_{\mathbf{k}} : K \textit{ is a compact operator}\}]^{1/n}. \quad \square$$

Lemma 1·7. *Spectral Decomposition*

Suppose $M : C^{\mathbf{k}}(X,Y) \to C^{\mathbf{k}}(X,Y)$ is a bounded linear operator with $\sigma(M) = \sigma_1 \cup \sigma_2$, where σ_1 consists of finitely many eigenvalues, each of finite multiplicity. Then there is a decomposition $C^{\mathbf{k}}(X,Y) = W^1 + W^2$ such that $\sigma(M|_{W^i}) = \sigma_i$, for $i = 1, 2$.

Proof. See Kato [24], page 178. $\square$

We now introduce the type of dynamical system we will be interested in.

Definition 1·2. *Let X be a d-dimensional partitioned manifold with partition $\mathcal{P} = \{P_1, \ldots, P_n\}$. Let us fix the regularity index $\mathbf{r} = (r, \delta)$, where $1 = (1,0) < \mathbf{r} = (r,\delta) < \infty = (\infty, 0)$. We say that $T : X \to X$ is a* **piecewise $C^{\mathbf{r}}$ expanding Markov map** *if*

1. *T is surjective,*
2. *$T(P_i)$ is a union of elements of $\mathcal{P}$, for each $1 \le i \le n$,*
3. *$T|_{P_i}$ is continuous for each $1 \le i \le n$,*
4. *$T|_{P_i}$ is $C^{\mathbf{r}}$ (except on the boundary ∂P_i, where we only demand that the appropriate directional derivatives exist) for each $1 \le i \le n$,*
5. *T is expanding. That is, there exists $\gamma < 1$ such that for all $1 \le i \le n$ and all $x \in int(P_i)$ we have*

$$|D_x T(v)|_d \ge \gamma^{-1} |v|_d \qquad \textit{for all } v \in \tau_x X.$$

Remarks.

(a) Condition 5 means that γ is a *contraction constant* for the family $\{T_j\}$ of local inverse branches of T. That is, $\|DT_j\|_\infty \le \gamma < 1$ for each inverse branch T_j.

(b) We will also be interested (see in particular §1·7 and §1·8) in piecewise $C^{\mathbf{r}}$ *eventually* expanding Markov maps. Such maps T satisfy properties 1–4 of the above definition, and further satisfy:

5′. T is eventually expanding. That is, there exists $\gamma < 1$ and $m \geq 1$ such that for all $1 \leq i \leq n$ and all $x \in \mathrm{int}(P_i)$ we have

$$|D_x T^m(v)|_d \geq \gamma^{-1}|v|_d \qquad \text{for all } v \in \tau_x X.$$

We note that a (conformal) change of norm will transform such a map into an expanding map.

Throughout §1·2, §1·3, §1·4 and §1·5 we will consider the map T (and its regularity index $\mathbf{r} = (r, \delta)$) as being fixed. We will define certain bounded linear operators in terms of T. These operators will leave invariant certain $C^{\mathbf{k}}$ spaces, where $\mathbf{k} \leq \mathbf{r} - 1 = (r-1, \delta)$.

In §1·6 we will consider conjugacies between such maps, with the further assumption that they are *conformal* (see Definition 1·8). Up until §1·6, however, this further assumption is not necessary.

The following important result uses the fact that $\mathbf{r} > 1 = (1, 0)$.

Lemma 1·8. *Let $T : X \to X$ be a piecewise $C^{\mathbf{r}}$, $\mathbf{r} > 1$, expanding Markov map of a partitioned manifold. There exists a T-invariant Borel probability measure m which is equivalent to Lebesgue measure.*

Proof. See Mañé [36], page 172. □

Remarks.

1. If T is topologically transitive then the measure m is unique.
2. m gives positive measure to all non-empty open subsets of X.

For $p > 0$, let $L^p(X, Y)$ denote the space of all Lebesgue-measurable functions $w : X \to Y$ for which

$$\|w\|_{L^p} = \int |w(x)|_Y^p \, dm(x) < \infty.$$

As usual we will consider $\|\cdot\|_{L^p}$ as a norm on the space $L^p(X,Y)$, though strictly speaking it is only a pseudo-norm. We note that each $L^p(X,Y)$ is a Banach space with respect to the L^p norm.

The following lemma will be used in the proof of Theorem 1·37.

Lemma 1·9. *Suppose $\phi_1,\ldots,\phi_d \in L^2(X,\mathbf{R})$. Then the product $\phi_1 \ldots \phi_d \in L^{2/d}(X,\mathbf{R})$.*

Proof. This is a consequence of the well-known Hölder inequality (see page 101 of Cohn [13], for example). □

If $M : L^p(X,Y) \to L^p(X,Y)$ is a bounded linear operator, then we define the operator norm $|\!|\!|\cdot|\!|\!|_{L^p}$ by

$$|\!|\!| M |\!|\!|_{L^p} = \sup\{\|Mw\|_{L^p} : w \in L^p(X,Y) \text{ with } \|w\|_{L^p} = 1\}.$$

The following lemma is a consequence of the well-known Fatou's Lemma.

Lemma 1·10. *Suppose for all $n \geq 0$ we have*

(i) $v_n \in L^1(X,\mathbf{R})$,

(ii) $v_n(x) \geq 0$,

(iii) $\liminf_{n\to\infty} \|v_n\|_{L^1} < \infty$,

(iv) $\lim_{n\to\infty} v_n(x)$ *exists for a.e.* $x \in X$.

Then

$$\int \lim_{n\to\infty} v_n(x)dm(x) \leq \liminf_{n\to\infty} \|v_n\|_{L^1}.$$

Proof. Fatou's Lemma is a classical result of measure theory. See page 72 of Cohn [13], for example. □

Note that $\|\cdot\|_{L^1}$ also defines a norm on $C^{\mathbf{k}}(X,Y)$ for any $\mathbf{k} \geq 0$, though the space is not a Banach space with respect to this norm. It is well-known that $C^{\mathbf{k}}(X,Y)$ is L^1-dense in $L^1(X,Y)$.

The following elementary lemma simply states that a derivative of a continuous function cannot have a *removable* point of discontinuity.

Lemma 1·11. *Suppose $w = Dh : X \to M(d)$ is the (almost everywhere defined) derivative of a piecewise continuous, (almost everywhere) differentiable map $h : X \to X$. Suppose there exists some $w' \in C^{\mathbf{k}}(X, M(d))$, $\mathbf{k} \geq 0$, such that $w' = w$ a.e. (m).*

Then $w = w'$ everywhere. So in particular w is defined everywhere, and $w \in C^{\mathbf{k}}(X, M(d))$.

Proof. See Spivak [67], page 178. □

Given our piecewise $C^{\mathbf{r}}$ expanding Markov map T, we define the function $f : X \to \mathbf{R}^+$ by

$$f(x) = 1/|\det D_x T|.$$

Note that $f \in C^{\mathbf{r}-1}(X, \mathbf{R}^+) = C^{(r-1,\delta)}(X, \mathbf{R}^+)$, where $\mathbf{r} = (r, \delta)$ is the regularity index of T (see Definition 1·2).

Definition 1·3. *The* **Ruelle-Perron-Frobenius operator** $\bar{L}_f : F(X, \mathbf{C}) \to F(X, \mathbf{C})$ *is the bounded linear operator defined by the formula*

$$\bar{L}_f(w)(x) = \sum_{Ty=x} f(y)w(y). \tag{1·1}$$

We have the following.

Lemma 1·12.

(i) If $\mathbf{k} \leq \mathbf{r} - \mathbf{1}$ then $C^{\mathbf{k}}(X, \mathbf{C})$ is $\bar{L}_f$-invariant.

(ii) $L^1(X, \mathbf{C})$ is $\bar{L}_f$-invariant. □

Let $h_\mu(T)$ denote the entropy of T with respect to a T-invariant Borel probability measure μ. Recall that the **pressure** $P(w)$ of a continuous function $w : X \to \mathbf{R}$ is defined as

$$P(w) = \sup\left\{ h_\mu(T) + \int w \, d\mu : \mu \text{ is a } T\text{-invariant probability measure} \right\}.$$

Further details on entropy and pressure can be found in Walters [72].

The following result is well known, though it is most often stated in the case where T is topologically mixing (in which case the statement is somewhat simpler). However we do not require our expanding map T to be mixing.

Theorem 1·13. *(Ruelle, [61]) Ruelle-Perron-Frobenius Theorem*

(i) If $0 < \mathbf{k} \leq \mathbf{r} - 1$ then $\bar{L}_f : C^{\mathbf{k}}(X, \mathbf{C}) \to C^{\mathbf{k}}(X, \mathbf{C})$ has spectral radius 1. There are a finite (and non-zero) number of eigenvalues of modulus 1. Each of these eigenvalues has a finite dimensional eigenspace. In particular, the number 1 is an eigenvalue, and it has a strictly positive eigenfunction $h \in C^{\mathbf{k}}(X, \mathbf{C})$.

(ii) The measure m (see Lemma 1·8) satisfies

$$\int \bar{L}_f w \, dm = \int w \, dm \qquad \text{for all } w \in L^1(X, \mathbf{C}).$$

(iii) The topological pressure $P(\log f) = 0$.

(iv) Let $V \subset C^{\mathbf{k}}(X, \mathbf{C})$ denote the span of the eigenspaces corresponding to the eigenvalues of maximum modulus. If $w \in C^0(X, \mathbf{C})$ then there exists $v \in V$ such that

$$\bar{L}_f^n w \to v \qquad \text{as } n \to \infty$$

in the supremum norm. □

In what follows, it will be convenient to *normalise* the operator $\bar{L}_f$ so as to define a new operator $\bar{L} : F(X, \mathbf{C}) \to F(X, \mathbf{C})$. We define the function $g : X \to \mathbf{R}^+$ by

$$g(x) = \frac{f(x)h(x)}{h(T(x))} \qquad \text{for all } x \in X.$$

Note that $g \in C^{\mathbf{r}-1}(X, \mathbf{R}^+) = C^{(r-1,\delta)}(X, \mathbf{R}^+)$.

Definition 1·4.

Define the bounded linear operator $\bar{L} : F(X, \mathbf{C}) \to F(X, \mathbf{C})$ by the formula

$$\bar{L}(w)(x) = \sum_{Ty=x} g(y)w(y). \tag{1·2}$$

We call $\bar{L}$ *the* **normalised Ruelle-Perron-Frobenius operator**.

We have the following analogues of Lemma 1·12 and Theorem 1·13.

Lemma 1·14.

(i) If $\mathbf{k} \le \mathbf{r} - 1$ *then* $C^{\mathbf{k}}(X, \mathbf{C})$ *is* $\bar{L}$*-invariant.*

(ii) $L^1(X, \mathbf{C})$ *is* $\bar{L}$*-invariant.* □

Theorem 1·15. *(Ruelle, [61]) Normalised Ruelle-Perron-Frobenius Theorem*

(i) If $0 < \mathbf{k} \le \mathbf{r} - 1$ *then* $\bar{L} : C^{\mathbf{k}}(X, \mathbf{C}) \to C^{\mathbf{k}}(X, \mathbf{C})$ *has spectral radius 1. There are a finite (and non-zero) number of eigenvalues of modulus 1. Each of these eigenvalues has a finite dimensional eigenspace. In particular, the number 1 is an eigenvalue, with the constant function 1 as an eigenfunction.*

(ii) The measure m *(see Lemma 1·8) satisfies*

$$\int \bar{L}w \, dm = \int w \, dm \qquad \text{for all } w \in L^1(X, \mathbf{C}).$$

(iii) The topological pressure $P(\log g) = 0$.

(iv) Let $V \subset C^{\mathbf{k}}(X, \mathbf{C})$ *denote the span of the eigenspaces corresponding to the eigenvalues of maximum modulus. If* $w \in C^0(X, \mathbf{C})$ *then there exists* $v \in V$ *such that*

$$\bar{L}^n w \to v \qquad \text{as } n \to \infty$$

in the supremum norm. □

We now want to define a bounded linear operator acting on functions w taking values in $\mathbf{C}^{d^2}$. We simply generalise the definition of $\bar{L}$ as follows.

Definition 1·5. *The bounded linear operator* $L : F(X, \mathbf{C}^{d^2}) \to F(X, \mathbf{C}^{d^2})$ *is defined by*

$$L(w)(x) = \sum_{Ty=x} g(y) w(y).$$

We call this the **vector operator**.

The analogue of Lemma 1·12 holds. That is,

Lemma 1·16.

(i) If $\mathbf{k} \le \mathbf{r} - 1$ *then* $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ *is* L*-invariant.*

(ii) $L^1(X, \mathbf{C}^{d^2})$ *is* L*-invariant.* $\square$

Section 1·3. The Matrix Operator.

In this section we introduce an operator L_θ defined in terms of a $C^{\mathbf{k}}$ unitary matrix-valued function θ. This operator was first defined, in the context of Hölder functions on a subshift of finite type, in Parry & Pollicott [49]. We investigate the spectral properties of L_θ acting on the space $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$. The results in this direction are similar in spirit to those of Ruelle [62] and Tangerman [68], and will be the key to the cocycle rigidity results of §1·4.

Suppose that $\theta \in C^{\mathbf{k}}(X, U(d^2))$ for some $0 < \mathbf{k} \le \mathbf{r} - 1$.

An immediate consequence of Lemma 1·4 (ii) is that

Lemma 1·17. *If* $\theta \in C^{\mathbf{k}}(X, U(d^2))$ *then* $\|\theta\|_\infty = 1$. $\square$

Definition 1·6. *We define the* **matrix operator** $L_\theta : F(X, \mathbf{C}^{d^2}) \to F(X, \mathbf{C}^{d^2})$ *by*

$$L_\theta(w)(x) = \sum_{Ty=x} g(y)\theta(y)w(y). \tag{1·3}$$

Note that

$$L_\theta w = L(\theta w), \tag{1·4}$$

where we define $(\theta w)(x) = \theta(x)w(x)$ for all $x \in X$.

Again, the analogue of Lemma 1·12 holds. That is,

Lemma 1·18.

(i) If $\mathbf{k} \leq \mathbf{r} - 1$ *then* $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ *is* L_θ*-invariant.*

(ii) $L^1(X, \mathbf{C}^{d^2})$ *is* L_θ*-invariant.* □

The following lemma shows that L_θ is well-defined on the almost-everywhere equivalence classes of $L^1(X, \mathbf{C}^{d^2})$.

Lemma 1·19. *Suppose* $v, v' \in L^1(X, \mathbf{C}^{d^2})$ *satisfy* $v = v'$ a.e. (m). *Then*

$$L_\theta v = L_\theta v' \quad \text{a.e. } (m).$$

Proof. If v is measurable then so is $L_\theta v$. Moreover, if $f, g : X \to \mathbf{R}^+$ are as in §1·2 then we have

$$\|L_\theta v\|_{L^1} \leq C \|\bar{L}_f v\|_{L^1} = C \|v\|_{L^1} \ ,$$

where $C = \sup_{x \in X} |g(x)/f(x)| < \infty$.

Thus if $\|v\|_{L^1} = 0$ then $\|L_\theta v\|_{L^1} = 0$. The result follows. □

We have the following lemma relating L_θ and $\bar{L}$.

Lemma 1·20. *For any* $w \in F(X, \mathbf{C}^{d^2})$, $x \in X$ *and* $n \geq 0$ *we have*

$$|L_\theta^n w(x)|_{d^2} \leq \bar{L}^n(|w(x)|_{d^2}).$$

Proof.

$$\begin{aligned}
|L_\theta^n w(x)|_{d^2} &= \left| \sum_{T^n y = x} g(T^{n-1}y) \ldots g(Ty) g(y) \theta(T^{n-1}y) \ldots \theta(Ty)\theta(y) w(y) \right|_{d^2} \\
&\leq \sum_{T^n y = x} g(T^{n-1}y) \ldots g(Ty) g(y) \ \left|\theta(T^{n-1}y) \ldots \theta(Ty)\theta(y)\right|_{d^2} |w(y)|_{d^2} \\
&= \sum_{T^n y = x} g(T^{n-1}y) \ldots g(Ty) g(y) \ |w(y)|_{d^2} \quad \text{by Lemma 1·4 (ii)} \\
&= \bar{L}^n(|w(x)|_{d^2}). \quad \square
\end{aligned}$$

The following simple corollary plays a crucial rôle in the proof of Proposition 1·28.

Corollary 1·21. $|||L_\theta^n|||_{L^1} \le 1$ *for all* $n \ge 0$.

Proof. For any $w \in L^1(X, \mathbb{C}^{d^2})$ we have

$$\begin{aligned}
\|L_\theta^n w\|_{L^1} &= \int |L_\theta^n w(x)|_{d^2} dm(x) \\
&\le \int \bar{L}^n(|w(x)|_{d^2}) dm(x) \quad \text{by Lemma 1·20} \\
&= \int |w(x)|_{d^2} dm(x) \quad \text{by Theorem 1·15 (ii)} \\
&= \|w\|_{L^1} .
\end{aligned}$$

Hence $|||L_\theta^n|||_{L^1} \le 1$. □

The following three technical results give information on how the operator L_θ acts on (piecewise) $C^{\mathbf{k}}$ functions.

Lemma 1·22. *Suppose* $w \in C^{\mathbf{k}}(X, \mathbb{C}^{d^2})$ *where* $0 < \mathbf{k} = (k, \epsilon) \le \mathbf{r} - 1$. *Then for any* $\gamma_0 \in (\gamma, 1)$ *there exists* $C_1 > 0$ *such that for all* $n \ge 0$*:*

$$\| L_\theta^n w \|_k \le C_1 \sum_{j=0}^{k} \| D^j w \|_\infty \gamma_0^{nj}.$$

Proof. Tangerman [68] noted the analogous result for the vector operator L (see Definition 1·5), and this same result was mentioned in Pollicott [53], [54], and Ruelle [61]. However, there does not seem to be a full proof anywhere in the literature. For this reason, as well as to demonstrate why the result remains true for the operator L_θ, we give the proof of this lemma in full. Since the proof is rather long and technical, we present it as Appendix A. □

Lemma 1·23. *Suppose* $w \in C^{\mathbf{k}}(X, \mathbb{C}^{d^2})$ *where* $0 < \mathbf{k} = (k, \epsilon) \le \mathbf{r} - 1$. *Then for any* $\gamma_0 \in (\gamma, 1)$ *there exists* $C_2 > 0$ *such that for all* $n \ge 0$*:*

$$\| D^k(L_\theta^n w) \|_\epsilon \le C_2 \left[\sum_{j=0}^{k} \left(\| D^j w \|_\infty \gamma_0^{nj} \right) + \|D^k w\|_\epsilon \gamma_0^{n(k+\epsilon)} \right].$$

Proof. This lemma is a generalisation of the so-called 'basic inequality' $\|L_\theta^n w\|_\epsilon \le C_2(\|w\|_\infty + \gamma_0^{n\epsilon}\|w\|_\epsilon)$ which is proved (in a slightly different context) on page 20 of Parry

& Pollicott [50]. The generalisation to higher derivatives uses in addition the ideas from the proof of Lemma 1·22 (see Appendix A). $\square$

Proposition 1·24. *Suppose $w \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ where $0 < \mathbf{k} \leq \mathbf{r} - 1$. Then for any $\gamma_0 \in (\gamma, 1)$ there exists $C > 0$ such that for all $n \geq 0$:*

$$\| L_\theta^n w \|_{\mathbf{k}} \leq C \left[\sum_{j=0}^{k} \left(\| D^j w \|_\infty \gamma_0^{nj} \right) + \|D^k w\|_\epsilon \gamma_0^{n|\mathbf{k}|} \right].$$

Proof. This result follows immediately from Lemmas 1·22 and 1·23:

$$\begin{aligned} \|L_\theta^n w\|_{\mathbf{k}} &= \|L_\theta^n w\|_k + \|D^k(L_\theta^n w)\|_\epsilon \\ &\leq (C_1 + C_2) \sum_{j=0}^{k} \| D^j w \|_\infty \gamma_0^{nj} + C_2 \|D^k w\|_\epsilon \gamma_0^{n(k+\epsilon)} \\ &\leq C \left[\sum_{j=0}^{k} \left(\| D^j w \|_\infty \gamma_0^{nj} \right) + \|D^k w\|_\epsilon \gamma_0^{n|\mathbf{k}|} \right] \end{aligned}$$

where $C = C_1 + C_2$. $\square$

Corollary 1·25. *If $0 < \mathbf{k} \leq \mathbf{r} - 1$ then there exists $C > 0$ such that $\| L_\theta^n \|_{\mathbf{k}} \leq C$ for all $n \geq 0$.*

Proof. Suppose $w \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$. Then for all $n \geq 0$ we have

$$\begin{aligned} \| L_\theta^n w \|_{\mathbf{k}} &\leq C \left[\sum_{j=0}^{k} \left(\| D^j w \|_\infty \gamma_0^{nj} \right) + \|D^k w\|_\epsilon \gamma_0^{n|\mathbf{k}|} \right] && \text{by Proposition 1·24} \\ &\leq C \left[\sum_{j=0}^{k} \| D^j w \|_\infty + \|D^k w\|_\epsilon \right] && \text{since } \gamma_0 < 1 \\ &= C \| w \|_{\mathbf{k}} . \end{aligned}$$

It follows that $\| L_\theta^n \|_{\mathbf{k}} \leq C$. $\square$

So we see from Corollaries 1·21 and 1·25 that iterates of L_θ are uniformly bounded in both the L^1 and the $C^{\mathbf{k}}$ norm.

We can use Proposition 1·24, together with Lemma 1·6, to prove the following.

Proposition 1·26.

Suppose $\gamma_0 \in (\gamma, 1)$, *and consider the operator* $L_\theta : C^{\mathbf{k}}(X, \mathbf{C}^{d^2}) \to C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$. *Then* $\sigma(L_\theta) \cap \{z \in \mathbf{C} : |z| \geq \gamma_0^{|\mathbf{k}|}\}$ *consists of a finite number of isolated eigenvalues, each of which has finite multiplicity.*

Proof. The details of this proof, using Taylor's Theorem, are given in Ruelle [62]. A very readable account of this method is given in Pollicott [54]. A slightly different approach, using Fourier approximations, is outlined in Pollicott [53] and Tangerman [68]. □

Theorem 1·27. *The space* $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ *can be decomposed as*

$$C^{\mathbf{k}}(X, \mathbf{C}^{d^2}) = W^1 + W^2,$$

where

(i) W^1 and W^2 *are both* L_θ*-invariant,*

(ii) W^1 *is finite dimensional,*

(iii) $L_\theta|_{W^2}$ *has spectral radius no more than* $\gamma_0^{|\mathbf{k}|}$, *where* $\gamma_0 \in (\gamma, 1)$.

Proof. This is immediate from Proposition 1·26 and Lemma 1·7. □

Section 1·4. The Matrix-valued Cocycle Equation.

In this section we study the Livsic cocycle equation

$$w(Tx)u(x) = v(x)w(x) \tag{1·5}$$

where u, v are (piecewise) $C^{\mathbf{k}}$ unitary matrix-valued functions, w is an L^1 matrix-valued function, and $T : X \to X$ is our piecewise $C^{\mathbf{r}}$ expanding Markov map of a partitioned manifold.

We show (Theorem 1·30) that an integrable matrix-valued function w satisfying (1·5) almost everywhere has a $C^{\mathbf{k}}$ version. That is, there exists a (piecewise) $C^{\mathbf{k}}$ function w' which satisfies (1·5) everywhere, and is equal to w almost everywhere.

This type of cocycle rigidity result was first obtained by Livsic [31], [32], who studied (1·5) for an Anosov diffeomorphism T and real-valued functions u, v, w. Further results in this direction were obtained by de la Llave, Marco & Moriyón [34]. We discuss real-valued cocycle equations further in §1·5, where we give a new proof of a result essentially contained in [34]. An adapted version of the original proof is contained in Appendix B. These two proofs rely (respectively) on the well-developed Ruelle-Perron-Frobenius theory for real-valued functions, and the fact that $\mathbf{R}$ is an abelian group under addition. Consequently neither proof is directly applicable to the matrix-valued cocycle equation.

Instead, our method of proof in this section is analogous to that used by Parry & Pollicott [49]. We transform equation (1·5) into a matrix-vector equation, and then into an operator equation involving L_θ. The details of these transformations are found in the proofs of Corollary 1·29 and Theorem 1·30. The main substance of this section is contained in the following proposition.

Proposition 1·28. *Suppose the matrix operator L_θ is defined (see Definition 1·6) in terms of a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map $T : X \to X$ of a partitioned manifold and a function $\theta \in C^{\mathbf{k}}(X, U(d^2))$, where $0 < \mathbf{k} \leq \mathbf{r} - 1$.*

Suppose $w \in L^1(X, \mathbf{C}^{d^2})$ satisfies

$$L_\theta w = w \quad \text{a.e. } (m).$$

Then there exists $w' \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ such that

$$(i) \quad w' = w \text{ a.e. } (m),$$

$$(ii) \quad L_\theta w' = w' \quad \textit{everywhere.}$$

Proof. (i) Since $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ is L^1-dense in $L^1(X, \mathbf{C}^{d^2})$, for every $j \geq 1$ we can choose some $w_j \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ such that

$$\int |(w - w_j)(x)|_{d^2} dm(x) < 1/j.$$

For each $j \geq 1$, consider the sequence $(L_\theta^n w_j)_{n=0}^\infty$.

By Theorem 1·27 we can write $w_j = w_j^1 + w_j^2$, where $w_j^1 \in W^1$ and $w_j^2 \in W^2$. The spaces W^1 and W^2 are both L_θ-invariant. The space W^1 is finite dimensional, while the spectral radius of $L_\theta|_{W^2}$ is less than or equal to $\gamma_0^{|\mathbf{k}|} < 1$.

So

$$L_\theta^n(w_j - w_j^1) = L_\theta^n w_j^2 \ ,$$

and thus

$$\| L_\theta^n w_j - L_\theta^n w_j^1 \|_{\mathbf{k}} = \| L_\theta^n w_j^2 \|_{\mathbf{k}} \ .$$

Now

$$\begin{aligned} \lim_{n\to\infty} |||(L_\theta|_{W^2})^n|||_{\mathbf{k}} &= \lim_{n\to\infty} \left[|||(L_\theta|_{W^2})^n|||_{\mathbf{k}}^{1/n} \right]^n \\ &= \lim_{n\to\infty} \left(\gamma_0^{|\mathbf{k}|}\right)^n \qquad \text{by Lemma 1·5} \\ &= 0 \qquad \text{since } 0 < \gamma_0^{|\mathbf{k}|} < 1. \end{aligned}$$

So for all $j \geq 1$ we have

$$\begin{aligned} \lim_{n\to\infty} \| L_\theta^n w_j - L_\theta^n w_j^1 \|_{\mathbf{k}} &= \lim_{n\to\infty} \| L_\theta^n w_j^2 \|_{\mathbf{k}} \\ &\leq \lim_{n\to\infty} |||(L_\theta|_{W^2})^n|||_{\mathbf{k}} \| w_j^2 \|_{\mathbf{k}} \\ &= 0. \end{aligned} \tag{1·6}$$

Now by Corollary 1·25 we know there exists some $C > 0$ such that

$$\| L_\theta^n w_j^1 \|_{\mathbf{k}} \leq C \| w_j^1 \|_{\mathbf{k}} \ \text{ for all } n \geq 0.$$

So the sequence $(L_\theta^n w_j^1)_{n=0}^\infty$ belongs to the set $\{v \in W^1 : \| v \|_{\mathsf{k}} \leq C \| w_j^1 \|_{\mathsf{k}}\}$. But since W^1 is finite dimensional then this set is compact, so that the sequence $(L_\theta^n w_j^1)_{n=0}^\infty$ has a C^{k} limit point, $w_j^\star$ say. Note that

$$w_j^\star \in W^1 \subset C^{\mathsf{k}}(X, \mathbf{C}^{d^2}). \tag{1·7}$$

In other words, there is a subsequence $(a_n)_{n=1}^\infty$ of the natural numbers such that

$$\lim_{n \to \infty} L_\theta^{a_n} w_j^1 = w_j^\star \text{ in the } C^{\mathsf{k}} \text{ norm.} \tag{1·8}$$

Then comparing equations (1·6) and (1·8) gives

$$\lim_{n \to \infty} L_\theta^{a_n} w_j = w_j^\star \text{ in the } C^{\mathsf{k}} \text{ norm.} \tag{1·9}$$

We now want to look at the behaviour of the sequence $(w_j^\star)_{j=1}^\infty$.

For almost every $x \in X$ we have that

$$\begin{aligned} |(w - w_j^\star)(x) - L_\theta^{a_n}(w - w_j)(x)|_{d^2} &= |L_\theta^{a_n} w_j(x) - w_j^\star(x)|_{d^2} \\ &\leq \| L_\theta^{a_n} w_j - w_j^\star \|_\infty \\ &\leq \| L_\theta^{a_n} w_j - w_j^\star \|_{\mathsf{k}} \\ &\longrightarrow 0 \quad \text{as } n \to \infty \text{ , by (1·9).} \end{aligned}$$

In other words,

$$(w - w_j^\star)(x) = \lim_{n \to \infty} L_\theta^{a_n}(w - w_j)(x) \quad \text{a.e. } (m), \tag{1·10}$$

where the limit is with respect to the $| \cdot |_{d^2}$ norm.

We would like to use Lemma 1·10 with $v_n(x) := |L_\theta^{a_n}(w - w_j)(x)|_{d^2}$. First we check that the hypotheses of Lemma 1·10 are satisfied.

Note that $v_n \in L^1(X, \mathbf{R})$ since $w \in L^1(X, \mathbf{C}^{d^2})$, $w_j \in C^{\mathsf{k}}(X, \mathbf{C}^{d^2}) \subset L^1(X, \mathbf{C}^{d^2})$, and $L^1(X, \mathbf{C}^{d^2})$ is L_θ-invariant (by Lemma 1·18 (ii)).

Clearly we also have that $v_n(x) \geq 0$.

By Corollary 1·21 we know that $\liminf_{n \to \infty} \|v_n\|_{L^1} < \infty$.

By the equality (1·10) we know that $\lim_{n \to \infty} v_n(x)$ exists almost everywhere.

So all the hypotheses of Lemma 1·10 are satisfied. Therefore

$$
\begin{aligned}
\int |(w - w_j^\star)(x)|_{d^2}\, dm(x) &= \int |\lim_{n\to\infty} L_\theta^{a_n}(w - w_j)(x)|_{d^2}\, dm(x) \quad \text{by } (1\cdot 10)\\
&= \int \lim_{n\to\infty} |L_\theta^{a_n}(w - w_j)(x)|_{d^2}\, dm(x)\\
&\le \liminf_{n\to\infty} \int |L_\theta^{a_n}(w - w_j)(x)|_{d^2}\, dm(x) \quad \text{by Lemma 1·10}\\
&\le \liminf_{n\to\infty} \int |(w - w_j)(x)|_{d^2}\, dm(x) \quad \text{by Corollary 1·21}\\
&= \int |(w - w_j)(x)|_{d^2}\, dm(x)\\
&< 1/j \quad \text{by choice of } w_j.
\end{aligned}
\tag{1·11}
$$

But then

$$
\begin{aligned}
\int |w_i^\star(x) - w_j^\star(x)|_{d^2}\, dm(x) &\le \int |w_i^\star(x) - w(x)|_{d^2}\, dm(x) + \int |w(x) - w_j^\star(x)|_{d^2}\, dm(x)\\
&< 1/i + 1/j \quad \text{by } (1\cdot 11),
\end{aligned}
$$

so that $(w_j^\star)_{j=1}^\infty$ is a Cauchy sequence, and hence a convergent sequence, with respect to the L^1 norm.

By (1·7) we know that each $w_j^\star \in W^1$. Since W^1 is finite dimensional then all norms are equivalent, so that $(w_j^\star)_{j=1}^\infty$ is convergent with respect to *any* norm, in particular the $C^{\mathbf{k}}$ norm.

So the pointwise limit $w'(x) := \lim_{j\to\infty} w_j^\star(x)$ exists for all $x \in X$.

Moreover, since w' is the $C^{\mathbf{k}}$ limit of a sequence of $C^{\mathbf{k}}$ functions, then $w' \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$.

We also have that

$$
\begin{aligned}
\int |w(x) - w'(x)|_{d^2}\, dm(x) &\le \int |w(x) - w_j^\star(x)|_{d^2}\, dm(x) + \int |w_j^\star(x) - w'(x)|_{d^2}\, dm(x)\\
&\longrightarrow 0 \quad \text{as } j \to \infty.
\end{aligned}
$$

Thus

$$
\int |w(x) - w'(x)|_{d^2}\, dm(x) = 0,
$$

from which we deduce that

$$
w = w' \quad \text{a.e. } (m).
$$

This completes the proof of part (i).

(ii) Now since $w' = w$ a.e. (m), then

$$\begin{aligned} L_\theta w' &= L_\theta w \text{ a.e. } (m) \text{ , by Lemma 1·19} \\ &= w \text{ a.e. } (m) \text{ , by hypothesis} \\ &= w' \text{ a.e. } (m). \end{aligned}$$

But $w' \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$. Thus $L_\theta w' \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ as well, since $C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ is L_θ invariant (see Lemma 1·18 (i)).

So the two functions w' and $L_\theta w'$ are equal a.e. (m), and are also both $C^{\mathbf{k}}$.

Therefore they must be equal *everywhere*. □

Corollary 1·29. *Suppose $T : X \to X$ is a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map of a partitioned manifold.*

Suppose $\theta \in C^{\mathbf{k}}(X, U(d^2))$ for some $0 < \mathbf{k} \leq \mathbf{r} - 1$.

Suppose $w \in L^1(X, \mathbf{C}^{d^2})$ satisfies the matrix-vector cocycle equation

$$\theta(x)w(x) = w(Tx) \quad \text{a.e. } (m). \tag{1·12}$$

Then there exists $w' \in C^{\mathbf{k}}(X, \mathbf{C}^{d^2})$ such that

(i) $w' = w$ a.e. (m),

(ii) $\theta(x)w'(x) = w'(Tx)$ everywhere.

Proof Applying the vector operator L (see Definition 1·5) to equation (1·12) gives

$$L_\theta w = w \quad \text{a.e. } (m). \tag{1·13}$$

By Proposition 1·28 (i) we deduce the existence of a $C^{\mathbf{k}}$ function w' such that $w' = w$ a.e. (m). This proves part (i) of the Corollary.

In fact, by an argument in Parry [48], equation (1·13) is *equivalent* to equation (1·12).

By Proposition 1·28 (ii) we know that w' satisfies

$$L_\theta w' = w' \quad \text{everywhere.}$$

Therefore w' satisfies

$$\theta(x)w'(x) = w'(Tx) \quad \text{everywhere,}$$

as required. $\square$

Theorem 1·30. *Suppose $T : X \to X$ is a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map of a partitioned manifold.*

Suppose $u, v \in C^{\mathbf{k}}(X, U(d))$ for some $0 < \mathbf{k} \leq \mathbf{r} - 1$. Suppose $w \in L^1(X, M(d))$ satisfies the cocycle equation

$$w(Tx)u(x) = v(x)w(x) \quad \text{a.e. } (m), \tag{1·14}$$

where the multiplication on both sides is $d \times d$ matrix multiplication.

Then there exists $w' \in C^{\mathbf{k}}(X, M(d))$ such that

(i) $w' = w$ a.e. (m),

(ii) $w'(Tx)u(x) = v(x)w'(x)$ *everywhere.*

Proof Since $u(x)$ is unitary for every $x \in X$, its inverse is equal to its adjoint $u(x)^\star$. So we can right-multiply equation (1·14) by $u(x)^\star$ to obtain

$$w(Tx) = v(x)w(x)u(x)^\star \quad \text{a.e. } (m). \tag{1·15}$$

For every $x \in X$ let us introduce the linear map

$$\begin{aligned} \theta(x) &: M(d) \to M(d) \\ &: A \mapsto v(x)Au(x)^\star \end{aligned}$$

Now for any $x \in X$ and $A, B \in M(d)$ we have

$$\begin{aligned} < \theta(x)(A), \theta(x)(B) >_{d^2} &=< v(x)Au(x)^\star, v(x)Bu(x)^\star >_{d^2} \\ &= \text{Trace}\,[v(x)Au(x)^\star(v(x)Bu(x)^\star)^\star] \quad \text{by Lemma 1·1} \\ &= \text{Trace}\,[v(x)Au(x)^\star u(x)B^\star v(x)^\star] \\ &= \text{Trace}\,[v(x)AB^\star v(x)^\star] \quad \text{since } u(x) \text{ is unitary} \\ &= \text{Trace}\,[AB^\star] \quad \text{by Lemma 1·2} \\ &=< A, B >_{d^2} \quad . \end{aligned}$$

So by Lemma 1·3 (ii) we deduce that $\theta(x) : M(d) \to M(d)$ is a unitary map.

Identifying $M(d)$ with $\mathbf{C}^{d^2}$, we can think of $\theta(x)$ as a $d^2 \times d^2$ unitary *matrix* acting on $\mathbf{C}^{d^2}$.

By the same identification we can think of w as belonging to $L^1(X, \mathbf{C}^{d^2})$.

So equation (1·15) is equivalent to

$$w(Tx) = \theta(x)w(x) \quad \text{a.e. } (m), \tag{1·16}$$

where the multiplication on the right hand side is of a $d^2 \times d^2$ matrix and a $d^2 \times 1$ vector.

Moreover, since $u(x), v(x)$ have a piecewise C^k dependence on x then so does $\theta(x)$. So we have a map $\theta \in C^k(X, U(d^2))$.

So the problem is equivalent to that of Corollary 1·29. By Corollary 1·29 (i), we can deduce that there exists $w' \in C^k(X, \mathbf{C}^{d^2})$ such that $w' = w$ a.e. (m). But we have identified $\mathbf{C}^{d^2}$ with $M(d)$, so $w' \in C^k(X, M(d))$ as required.

Moreover, by Corollary 1·29 (ii) we know that

$$\theta(x)w'(x) = w'(Tx) \quad \text{everywhere.}$$

Thus

$$w'(Tx)u(x) = v(x)w'(x) \quad \text{everywhere.} \qquad \square$$

Theorem 1·31.

Suppose $T : X \to X$ is a piecewise C^∞ expanding Markov map of a partitioned manifold.

Suppose $u, v \in C^\infty(X, U(d))$. Suppose $w \in L^1(X, M(d))$ satisfies the cocycle equation

$$w(Tx)u(x) = v(x)w(x) \quad \text{a.e. } (m),$$

where the multiplication on both sides is $d \times d$ matrix multiplication.

Then there exists $w' \in C^\infty(X, M(d))$ such that

(i) $w' = w$ a.e. (m),

(ii) $w'(Tx)u(x) = v(x)w'(x)$ *everywhere.*

Proof By Theorem 1·30 we see that for any $k \geq 1$ there exists $w' \in C^{(k,0)}(X, M(d))$ such that $w = w'$ almost everywhere. Clearly w' is independent of k. Thus $w' \in C^{\infty}(X, M(d))$. $\square$

Section 1·5. The Real-valued Cocycle Equation.

In this section we outline some of the rigidity theory for the real-valued cocycle equation. This theory dates back to the work of Livsic [31], [32], who considered Hölder functions in the context of Anosov diffeomorphisms and subshifts of finite type.

De la Llave, Marco & Moriyón [34] proved a smooth cocycle rigidity theorem in the context of C^{∞} Anosov diffeomorphisms, and in Appendix B we adapt this proof to the case of piecewise $C^{\mathbf{r}}$ expanding Markov maps. In this section, however, we present a much shorter proof (Theorem 1·33) of the same result. Our method uses Ruelle-Perron-Frobenius operators, and is similar in spirit to the techniques of §1·4.

We begin by stating a classical result due essentially to Livsic [31].

Theorem 1·32. *(Livsic, [31]) Suppose* $T : X \to X$ *is a piecewise* $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, *expanding Markov map of a partitioned manifold.*

Suppose $\Phi \in C^{(0,\epsilon)}(X, \mathbf{R})$ *for some* $0 < \epsilon < 1$.

Suppose $W \in L^{\infty}(X, \mathbf{R})$ *satisfies the real-valued cocycle equation*

$$W(Tx) - W(x) = \Phi(x) \quad \text{a.e. } (m).$$

Then there exists $W' \in C^{(0,\epsilon)}(X, \mathbf{R})$ *such that*

(i) $W' = W$ a.e. (m),

(ii) $W'(Tx) - W'(x) = \Phi(x)$ *everywhere.* $\square$

The following result is a smooth generalisation of Theorem 1·32.

Theorem 1·33. *Suppose $T : X \to X$ is a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map of a partitioned manifold.*

Suppose $\Phi \in C^{\mathbf{k}}(X, \mathbf{R})$ for some $0 < \mathbf{k} = (k, \epsilon) \leq \mathbf{r} - 1$.

Suppose $W \in L^{\infty}(X, \mathbf{R})$ satisfies the real-valued cocycle equation

$$W(Tx) - W(x) = \Phi(x) \quad \text{a.e. } (m). \tag{1·17}$$

Then there exists $W' \in C^{\mathbf{k}}(X, \mathbf{R})$ such that

(i) $W' = W$ a.e. (m),

(ii) $W'(Tx) - W'(x) = \Phi(x)$ *everywhere.*

Proof. Since $W \in L^{\infty}(X, \mathbf{R})$, then by Theorem 1·32 there exists $W' \in C^{(0,\epsilon)}(X, \mathbf{R})$ such that $W = W'$ almost everywhere. So without loss of generality we will assume that $W \in C^{(0,\epsilon)}(X, \mathbf{R})$, and that the cocycle equation (1·17) holds for *all* $x \in X$.

If we define $w(x) = e^{W(x)}$ and $\phi(x) = e^{\Phi(x)}$, then (1·17) becomes

$$\phi(x)w(x) = w(Tx), \tag{1·18}$$

where $\phi \in C^{\mathbf{k}}(X, \mathbf{R}^+)$ and $w \in C^{(0,\epsilon)}(X, \mathbf{R}^+)$. Recall (see Definition 1·4) that the normalised Ruelle-Perron-Frobenius operator $\bar{L} : F(X, \mathbf{C}) \to F(X, \mathbf{C})$ is defined by

$$\bar{L}(u)(x) = \sum_{Ty=x} g(y)u(y),$$

where $g : X \to \mathbf{R}^+$ is a particular $C^{\mathbf{r}-1}$ function. By Theorem 1·15 we know that $\bar{L}1 = 1$.

We can define another operator $\bar{L}_\phi : F(X, \mathbf{C}) \to F(X, \mathbf{C})$ by

$$\bar{L}_\phi(u)(x) = \sum_{Ty=x} g(y)\phi(y)u(y).$$

By the same Ruelle-Perron-Frobenius result (see Ruelle [61] for details) as we used to obtain Theorem 1·13, we know that $\bar{L}_\phi$ has positive spectral radius λ, say. Also, there are finitely many eigenvalues of modulus λ, and each of these has a finite dimensional eigenspace. If V denotes the span of these eigenspaces, then there exists $v \in V$ such that

$$\lim_{n \to \infty} \frac{1}{\lambda^n} \bar{L}_\phi^n w = v \tag{1·19}$$

in the supremum norm.

Since w satisfies (1·18) for all $x \in X$, we obtain

$$\begin{aligned}
\bar{L}_\phi w(x) &= \sum_{Ty=x} g(y)\phi(y)w(y) \\
&= \sum_{Ty=x} g(y)w(Ty) \\
&= w(x) \sum_{Ty=x} g(y) \\
&= w(x)\bar{L}1 \\
&= w(x).
\end{aligned}$$

Consequently

$$\lim_{n\to\infty} \bar{L}_\phi^n w(x) = w(x) \quad \text{for all } x \in X. \tag{1·20}$$

In particular, $\lim_{n\to\infty} \bar{L}_\phi^n w(x)$ is finite and non-zero, so from (1·19) we see that $\lambda = 1$. Then comparing (1·19) and (1·20) gives $w = v \in V \subset C^{\mathbf{k}}(X, \mathbf{C})$. Hence $W \in C^{\mathbf{k}}(X, \mathbf{C})$. □

Section 1·6. Conjugacies between Piecewise Smooth Conformal Expanding Markov Maps

We have studied a piecewise C^r expanding Markov map $T : X \to X$ of a d-dimensional partitioned manifold X, and in §1·4 we considered a matrix-valued cocycle equation defined in terms of T. These matrices were of size $d \times d$, though their size need not have corresponded to the dimension of X. It is only in this section, where we consider a cocycle equation involving derivatives of maps between d-dimensional partitioned manifolds, that this correspondence becomes necessary. So as to apply our previous results, it will be necessary to make a further assumption on the type of maps we will study.

Definition 1·7. *Let $SO(d)$ denote the set of real-valued $d \times d$ unitary matrices with determinant 1. We call this the group of* **special orthogonal** *matrices.*

Definition 1·8. *A differentiable map $T : X \to X$ of a d-dimensional partitioned manifold is said to be* **conformal** *if there exist functions $U : X \to SO(d)$ and $a : X \to \mathbf{R}$ such that $D_x T = a(x)U(x)$ for all $x \in X$.*

We will consider almost everywhere differentiable conjugacies (see Definition 1·10) between piecewise C^r conformal expanding Markov maps, and prove (Theorem 1·37) that these conjugacies are themselves piecewise C^r.

Our method of proof is the following. By differentiating (almost everywhere) the conjugacy equation, we obtain a matrix cocycle equation. We split this matrix cocycle equation into two further cocycle equations, one involving real-valued functions, the other involving matrices with determinant ± 1. We will use the results of §1·4 and §1·5 to guarantee the piecewise C^{r-1} regularity of solutions to these equations. We then deduce the piecewise C^r regularity of our conjugacy.

We start by essentially re-stating part of Theorem 1·30. The only difference in the following version is that we have replaced the group $U(d)$ of unitary matrices by its subgroup $SO(d)$.

Theorem 1·34. *Suppose $T : X \to X$ is a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map of a partitioned manifold.*

Suppose $u, v \in C^{\mathbf{k}}(X, SO(d))$ for some $0 < \mathbf{k} \leq \mathbf{r} - 1$. Suppose $w \in L^1(X, M(d))$ satisfies the cocycle equation

$$w(Tx)u(x) = v(x)w(x) \quad \text{a.e. } (m),$$

where the multiplication on both sides is $d \times d$ matrix multiplication.

Then there exists $w' \in C^{\mathbf{k}}(X, M(d))$ such that $w' = w$ a.e. (m). □

Throughout this section X_1, X_2 will be d-dimensional partitioned manifolds with corresponding partitions $\mathcal{P}_1$, $\mathcal{P}_2$, while $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ will be piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} = (r, \delta) < \infty$, conformal expanding Markov maps with respect to these partitions. That is, each T_i is $C^{\mathbf{r}}$ except possibly on the boundary $\partial\mathcal{P}_i$, where we only demand that the appropriate directional derivatives of order r exist and are δ-Hölder.

Two principal examples of such maps are the following.

1. Any piecewise $C^{\mathbf{r}}$ expanding Markov map of a one-dimensional partitioned manifold (for instance the circle). Since every one-dimensional map is (trivially) conformal then conformality is no restriction.

2. A class of piecewise analytic conformal expanding Markov maps of S^n. Such maps are induced by Kleinian group actions. More details appear in §1·7 and §1·8.

The following definition of topological conjugacy differs from the usual one in that we only require it to be *piecewise* continuous with respect to the partition associated to our partitioned manifold.

Definition 1·9. *We say that $h : X_1 \to X_2$ is a* **topological conjugacy** *between T_1 and T_2 if*

(a) h is a bijection,

(b) h and h^{-1} are piecewise continuous (with respect to $\mathcal{P}_1$ and $\mathcal{P}_2$) on X_1 and X_2 respectively,

(c) $h(T_1(x)) = T_2(h(x))$ *for all* $x \in X_1$.

The next lemma is motivated by the definition that follows it.

Lemma 1·35. *Suppose* $h : X_1 \to X_2$ *is differentiable Lebesgue almost everywhere. Then the derivative* $Dh : X_1 \to M(d)$ *(defined almost everywhere) is Lebesgue measurable.*

Proof. In one dimension the derivative is the limit of a sequence of measurable functions, and hence itself measurable. In higher dimensions, just apply this argument to each of the directional derivatives. □

For our purposes the measurability of Dh will not be sufficient. We make the following definition.

Definition 1·10. *We say that* $h : X_1 \to X_2$ *is an* **almost everywhere differentiable conjugacy** *between* T_1 *and* T_2 *if*

(a) h *is a topological conjugacy between* T_1 *and* T_2,

(b) The set $\{x \in X_1 : h \text{ is differentiable at } x \text{ and } h^{-1} \text{ is differentiable at } h(x)\}$ *has full measure with respect to normalised Lebesgue measure* l *on* X_1,

(c) All partial derivatives of h *and* h^{-1} *are essentially bounded.*

Remark. We impose condition (b) to ensure that h is differentiable almost everywhere, and that this derivative is non-singular almost everywhere. Equivalently, we could require that $l(A) = 0$ if and only if $l(h(A)) = 0$ (a kind of *absolute continuity* requirement), and replace (b) with the condition that h and h^{-1} should both be differentiable almost everywhere.

Recall that in the one-dimensional case any Lipschitz map is absolutely continuous, and any absolutely continuous map is differentiable almost everywhere (see Royden [60], for example). The generalisation of absolute continuity to higher dimensions is somewhat problematic (in reality there are several possible generalisations), though by Rademacher's Theorem (see page 21 of Morgan [40]) we know that a (higher-dimensional) Lipschitz map is differentiable almost everywhere.

We want to prove that if h is an almost everywhere differentiable conjugacy between conformal expanding Markov maps, then in fact h is as smooth as the maps themselves. The first step in that direction is the following result, which immediately guarantees us Hölder regularity of any topological conjugacy. Note that here we do not require the expanding Markov maps to be conformal, and in fact they need only be (piecewise) C^1.

Lemma 1·36. *Let $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be piecewise $C^{\mathbf{r}}$, $\mathbf{r} \geq 1$, expanding Markov maps of the partitioned manifolds X_1, X_2. Suppose h is a topological conjugacy between T_1 and T_2. Then there exists $0 < \alpha < 1$ such that h is piecewise $C^{(0,\alpha)}$.*

Proof. The proof on page 599 of Katok & Hasselblatt [25] is for conjugacies between Anosov diffeomorphisms. This is easily adapted to the case of expanding maps. We remark that the Hölder exponent α depends on the expansion constants for T_1 and T_2. □

The following theorem is our main result of this chapter.

Theorem 1·37. *Suppose $1 < \mathbf{r} = (r, \delta) < \infty$. Let $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be piecewise $C^{\mathbf{r}}$ conformal expanding Markov maps of the partitioned manifolds X_1, X_2. Suppose that T_1 and T_2 both preserve (or both reverse) orientation. Suppose that $h : X_1 \to X_2$ is an almost everywhere differentiable conjugacy between T_1 and T_2.*

Then h is in fact a $C^{\mathbf{r}}$ conjugacy between T_1 and T_2. That is, h is $C^{\mathbf{r}}$ except on $\partial\mathcal{P}_1$, and h^{-1} is $C^{\mathbf{r}}$ except on $\partial\mathcal{P}_2$.

Proof. Since $h : X_1 \to X_2$ is a topological conjugacy between T_1 and T_2, we have

$$h(T_1 x) = T_2(hx) \quad \text{for all } x \in X_1. \tag{1·21}$$

Since T_1, T_2 are expanding, and h is differentiable almost everywhere, we can differentiate equation (1·21) to obtain

$$D_{T_1(x)}h.D_x T_1 = D_{h(x)}T_2.D_x h \quad \text{a.e. } (l). \tag{1·22}$$

Since T_1, T_2 are both conformal, we can write $D_x T_1 = a_1(x)U_1(x)$ and $D_x T_2 = a_2(x)U_2(x)$, where $a_i : X_i \to \mathbf{R}$ and $U_i : X_i \to SO(d)$. Then equation (1·22) becomes

$$D_{T_1(x)}h.a_1(x)U_1(x) = a_2(h(x))U_2(h(x)).D_x h \quad \text{a.e. } (l). \tag{1·23}$$

Let m be the T_1-invariant Borel probability measure which is equivalent to l (see Lemma 1·8). Then equation (1·23) gives

$$D_{T_1(x)}h.a_1(x)U_1(x) = a_2(h(x))U_2(h(x)).D_x h \quad \text{a.e. } (m). \tag{1·24}$$

Let $S(d) = \{d \times d \text{ matrices with determinant } \pm 1\}$. We want to split equation (1·24) into two equations, one involving real-valued functions, the other involving $S(d)$-valued functions.

Since $U_1(x), U_2(h(x)) \in SO(d)$ for all $x \in X_1$, taking the determinant of both sides of (1·24) gives

$$\det[D_{T_1(x)}h]a_1(x)^d = a_2(h(x))^d \det[D_x h] \quad \text{a.e. } (m). \tag{1·25}$$

Now since T_1, T_2 both preserve (or both reverse) orientation, then a_1, a_2 are both positive-valued (or both negative-valued) functions. In either case, equation (1·25) implies that $\det[D_{T_1(x)}h]$ and $\det[D_x h]$ have the same sign.

By taking the absolute value and the d^{th} root of both sides of (1·25), we obtain

$$|\det[D_{T_1(x)}h]|^{1/d}a_1(x) = a_2(h(x))|\det[D_x h]|^{1/d} \quad \text{a.e. } (m). \tag{1·26}$$

Substituting (1·26) into the matrix equation (1·24) gives

$$(|\det[D_{T_1(x)}h]|^{-1/d}D_{T_1(x)}h).U_1(x) = U_2(h(x)).(|\det[D_x h]|^{-1/d}D_x h) \quad \text{a.e. } (m). \tag{1·27}$$

To clarify equation (1·27) we define the function $w : X_1 \to S(d) \subset M(d)$ by

$$w(x) = |\det[D_x h]|^{-1/d}D_x h = |\det[D_{h(x)}h^{-1}]|^{1/d}D_x h \ ,$$

wherever this makes sense. The condition (b) in Definition 1·10 guarantees a set of full measure on which both h and h^{-1} are differentiable, so the function w is defined almost everywhere. By Lemma 1·35 we know that w is measurable, and we want to show further that w is L^1. Now $\det[D_{h(x)}h^{-1}]$ is the sum of $d!$ products, where each product has d factors, and each factor is a partial derivative of h^{-1}. In particular, since h is an almost everywhere differentiable conjugacy, each factor is an L^∞ function, and hence an L^2 function, of x. By Lemma 1·9 we deduce that each of the products is an $L^{2/d}$ function of x, and hence that $\det[D_{h(x)}h^{-1}]$ is an $L^{2/d}$ function of x. Thus $|\det[D_{h(x)}h^{-1}]|^{1/d}$ is an L^2 function of x. We also know (by condition (c) of Definition 1·10) that $D_x h$ is an L^∞ function, and hence an L^2 function, of x. Applying Hölder's inequality we deduce that w is an L^1 function of x. That is, $w \in L^1(X_1, M(d))$.

Then equation (1·27) becomes

$$w(T_1(x)).U_1(x) = U_2(h(x)).w(x) \quad \text{a.e. } (m). \tag{1·28}$$

This is a matrix-valued cocycle equation, where U_1, U_2 take values in $SO(d)$, and the solution w is integrable.

Returning to the real-valued equation (1·26), we note that at almost every $x \in X_1$ we have $\det[D_x h] \neq 0$ and $\det[D_{T_1(x)}h] \neq 0$, so at these points we can take logarithms of both sides to obtain

$$\log|\det[D_{T_1(x)}h]|^{1/d} - \log|\det[D_x h]|^{1/d} = \log a_2(h(x)) - \log a_1(x) \quad \text{a.e. } (m). \tag{1·29}$$

We can define $W : X_1 \to \mathbf{R}$ by

$$W(x) = \log|\det D_x h|^{1/d},$$

wherever this makes sense. Condition (b) of Definition 1·10 means that W is defined almost everywhere. Now condition (c) of Definition 1·10 means that the maps $x \mapsto |\det D_x h|$ and $x \mapsto |\det D_{h(x)}h^{-1}| = |\det D_x h|^{-1}$ are both essentially bounded. The same is true of the maps $x \mapsto |\det D_x h|^{1/d}$ and $x \mapsto |\det D_x h|^{-1/d}$. Noting also that $\log y \leq \max(y, y^{-1})$, we deduce that W is essentially bounded.

Then equation (1·29) becomes

$$W(T_1(x)) - W(x) = \log a_2(h(x)) - \log a_1(x) \quad \text{a.e. } (m). \tag{1·30}$$

This is a real-valued cocycle equation whose solution W is an L^∞ function.

Note that $D_x h = e^{W(x)} w(x)$ for almost all $x \in X_1$.

We can now prove that h is piecewise $C^{(r,\delta)}$ on X_1. The first step is to show that h is piecewise $C^{(0,\delta)}$.

By Lemma 1·36 we know h is piecewise $C^{(0,\alpha)}$ for some $0 < \alpha < 1$.

If $\alpha \geq \delta$ then this immediately implies that h is piecewise $C^{(0,\delta)}$.

If $\alpha < \delta$ then we will show that h is piecewise $C^{(1,\alpha)}$, and hence piecewise $C^{(0,\delta)}$.

Define $u : X_1 \to SO(d)$ by $u(x) = U_1(x)$ for all $x \in X_1$.

Define $v : X_1 \to SO(d)$ by $v(x) = U_2(h(x))$ for all $x \in X_1$.

So by equation (1·28) we have the matrix cocycle equation

$$w(T_1(x))u(x) = v(x)w(x) \quad \text{a.e. } (m), \tag{1·31}$$

where $w \in L^1(X_1, M(d))$. We would like to apply Theorem 1·34, but before doing so we must check the regularity of the $SO(d)$-valued functions u, v.

Certainly $u \in C^{(0,\alpha)}(X_1, SO(d))$, since

$$U_1 \in C^{(r-1,\delta)}(X_1, SO(d)) \subset C^{(0,\alpha)}(X_1, SO(d)).$$

Also $v \in C^{(0,\alpha)}(X_1, SO(d))$, since h is piecewise $C^{(0,\alpha)}$, and since

$$U_2 \in C^{(r-1,\delta)}(X_2, SO(d)) \subset C^{(0,\alpha)}(X_2, SO(d)).$$

By Theorem 1·34 (with $\mathbf{k} = (0, \alpha)$) we deduce that there exists $w' \in C^{(0,\alpha)}(X_1, M(d))$ such that $w = w'$ almost everywhere.

We can apply essentially the same argument to the real-valued cocycle equation (1·30) to deduce (using Theorem 1·33 with $\mathbf{k} = (0, \alpha)$) that there exists $W' \in C^{(0,\alpha)}(X_1, \mathbf{R})$ such that $W = W'$ almost everywhere.

So we know that $D_x h = e^{W(x)} w(x) = e^{W'(x)} w'(x)$ for almost all $x \in X_1$.

In other words, the piecewise $C^{(0,\alpha)}$ function $F : x \mapsto e^{W'(x)} w'(x)$ is almost everywhere equal to Dh.

But since Dh is a derivative of a continuous function, we can use Lemma 1·11 to deduce that $Dh = F$ everywhere.

So Dh is a piecewise $C^{(0,\alpha)}$ function.

Therefore h is piecewise $C^{(1,\alpha)}$. In particular this implies that h is piecewise $C^{(0,\delta)}$.

Therefore, irrespective of the Hölder constant α guaranteed by Lemma 1·36, we have shown that h is piecewise $C^{(0,\delta)}$.

We will now prove that h is piecewise $C^{(r,\delta)}$. The proof is by induction. Our inductive hypothesis is that h is piecewise $C^{(k,\delta)}$ for some $0 \le k \le r-1$, and we have just seen that this is true for $k = 0$.

We will show this implies that h is piecewise $C^{(k+1,\delta)}$.

The inductive hypothesis means that the functions u, v in the matrix-valued cocycle equation (1·31) are both piecewise $C^{(k,\delta)}$ (recall that v was defined in terms of h). By Theorem 1·34 we deduce that w is almost everywhere equal to a piecewise $C^{(k,\delta)}$ function. But we have seen that w is piecewise (Hölder) continuous. Therefore $w \in C^{(k,\delta)}(X_1, M(d))$.

A similar argument applied to equation (1·30), and using Theorem 1·33, shows that $W \in C^{(k,\delta)}(X_1, \mathbb{R})$.

Since $D_x h = e^{W(x)} w(x)$ then Dh is piecewise $C^{(k,\delta)}$, and hence h is piecewise $C^{(k+1,\delta)}$.

This concludes the inductive step, and so h is piecewise $C^{(r,\delta)}$.

An analogous argument shows that h^{-1} is piecewise $C^{(r,\delta)}$.

Therefore the theorem is proved. □

Theorem 1·38. *Let $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be piecewise C^∞ conformal expanding Markov maps of the partitioned manifolds X_1, X_2. Suppose that T_1 and T_2 both preserve (or both reverse) orientation. Suppose that $h : X_1 \to X_2$ is an almost everywhere differentiable conjugacy between T_1 and T_2.*

Then h is in fact a C^∞ conjugacy between T_1 and T_2. That is, h is C^∞ except on $\partial\mathcal{P}_1$, and h^{-1} is C^∞ except on $\partial\mathcal{P}_2$.

Proof. Immediate from Theorem 1·37. □

Section 1·7. A Class of Piecewise Analytic Conformal Expanding Markov Maps.

We will describe a class of piecewise analytic conformal expanding Markov maps defined on the n-sphere S^n. This class of maps satisfies the hypotheses of §1·6. Consequently we can apply Theorem 1·38 to conjugacies between them. Our exposition follows closely the paper of Bowen & Series [8], who studied these maps in the case $n = 1$. This class of maps arises naturally in the study of discrete subgroups of isometries of hyperbolic space, and there are links to the study of Riemannian manifolds of constant negative curvature.

In this section we will describe the construction of these maps in the simplest case $n = 1$, while in §1·8 we indicate how the construction is generalised to higher dimensions. In §1·9 we use these maps, together with Theorem 1·38, to give an alternative proof of a part of the well-known Mostow Rigidity Theorem.

A comprehensive reference to the background material of this section is Maskit [39]. Another good source is Ford [17].

Let $\mathbf{D} = \{x \in \mathbf{R}^2 : |x| < 1\}$ be the two-dimensional open unit disc. Its boundary $\partial\mathbf{D} = S^1 = \{x \in \mathbf{R}^2 : |x| = 1\}$ is the unit circle.

In what follows, it will be convenient to consider $\mathbf{D}$ and S^1 as subsets of the complex plane. So we have $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$ and $S^1 = \{z \in \mathbf{C} : |z| = 1\}$.

We consider $\mathbf{D}$ as the Poincaré disc model of two-dimensional hyperbolic space by giving it the hyperbolic metric

$$ds = \frac{2|dz|}{1 - |z|^2} \quad .$$

The boundary S^1 represents the circle at infinity. In this model of hyperbolic space, the geodesics of $\mathbf{D}$ are precisely the arcs of Euclidean circles orthogonal to S^1.

Definition 1·11. *Let $\mathcal{M}(\mathbf{D})$ denote the group of orientation-preserving isometries of $\mathbf{D}$ with respect to the hyperbolic metric. We call this the* **Möbius group**.

In what follows we will consider discrete subgroups of $\mathcal{M}(\mathbf{D})$, so first we will describe explicitly the elements of $\mathcal{M}(\mathbf{D})$.

Define the map $g : \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ of the extended complex plane to be the fractional linear transformation given by

$$g(z) = \frac{az + \bar{c}}{cz + \bar{a}} \quad , \qquad \text{where } a, c \in \mathbf{C} \text{ satisfy } |a|^2 - |c|^2 = 1. \tag{1·32}$$

The condition on the parameters a and c ensures that $g(S^1) = S^1$ and $g(\mathbf{D}) = \mathbf{D}$.

In fact, the elements of $\mathcal{M}(\mathbf{D})$ are precisely the restrictions to $\mathbf{D}$ of such maps g. By slight abuse of notation we will also let g denote this restriction.

Definition 1·12. *An element of $\mathcal{M}(\mathbf{D})$ is called* **parabolic** *if it has exactly one fixed point on S^1.*

The set $\{z \in \mathbf{C} : |g'(z)| > 1\} = \{z \in \mathbf{C} : |cz + \bar{a}| < 1\}$ is called the *isometric ball* of the map g. Note that the isometric ball is the empty set when g is the identity map, but that otherwise it is a Euclidean ball.

The set $\{z \in \mathbf{C} : |g'(z)| = 1\} = \{z \in \mathbf{C} : |cz + \bar{a}| = 1\}$ is called the *isometric circle* of g. When g is the identity map then the isometric circle is the whole complex plane, but otherwise it is a Euclidean circle.

The following lemma is elementary (see Ford [17] for a proof of part (i)).

Lemma 1·39. *Suppose $g \in \mathcal{M}(\mathbf{D})$ is not the identity map. Then*

(i) the isometric circle of g is orthogonal to the unit circle S^1,

(ii) g is conformal and complex analytic on its isometric ball,

(iii) $|g'(z)| > 1$ if and only if z lies inside the isometric ball of g. □

We now consider the restriction of g to $\overline{\mathbf{D}}$, and define the corresponding isometric ball (resp. circle).

We define $B(g) = \{z \in \overline{\mathbf{D}} : |g'(z)| < 1\} = \{z \in \overline{\mathbf{D}} : |cz + \bar{a}| < 1\}$,

and $S(g) = \{z \in \overline{\mathbf{D}} : |g'(z)| = 1\} = \{z \in \overline{\mathbf{D}} : |cz + \bar{a}| = 1\}$.

We will call $B(g)$ (resp. $S(g)$) the isometric ball (resp. circle) of g, even though it is really only a *part* of the isometric ball (resp. circle) defined previously.

We also define $A(g) = \{z \in S^1 : |cz + \bar{a}| \leq 1\}$ to be the isometric (closed) arc of g.

Definition 1·13. *A* **Fuchsian group** Γ *is a discrete subgroup of* $\mathcal{M}(\mathbf{D})$.

If $x \in \mathbf{D}$ then we define $\Gamma(x) = \{g(x) : g \in \Gamma\}$ to be the *orbit* of the point x under the group Γ.

The *limit set* $\Lambda \subset \overline{\mathbf{D}}$ of a Fuchsian group Γ is the set of accumulation points of orbits. It is easy to show (see Ford [17]) that in fact $\Lambda \subset S^1$.

We say that Γ is *elementary* if its limit set Λ consists of at most two points. If Γ is non-elementary then it can be shown (see Ford [17], for example) that its limit set Λ is a perfect set.

From now on we will assume that Γ is a finitely generated non-elementary Fuchsian group.

Definition 1·14. *We say that the open set* $\mathcal{R} \subset \mathbf{D}$ *is a* **fundamental domain** *for* Γ *if it satisfies:*

(a) Its boundary $\partial\mathcal{R}$ *has Lebesgue measure zero.*

(b) If $z_1, z_2 \in \mathcal{R}$ *then* $g(z_1) \neq z_2$ *for all* $g \in \Gamma$.

(c) If $z_1 \in \mathbf{D}$ *then there exists* $z_2 \in \overline{\mathcal{R}}$ *and* $g \in \Gamma$ *such that* $g(z_1) = z_2$.

Since $g(\mathcal{R}) \cap \mathcal{R} = \emptyset$ for every $g \in \Gamma$, there is plenty of freedom in our choice of the fundamental domain $\mathcal{R}$.

A natural choice of $\mathcal{R}$ is the subset of $\mathbf{D}$ which is exterior to all the (closed) isometric balls of the elements of Γ.

That is,

$$\mathcal{R} = \mathbf{D} \setminus \bigcup_{g \in \Gamma} \overline{B(g)}.$$

This choice of $\mathcal{R}$ is often called the Ford fundamental domain. This fundamental domain is a hyperbolic polygon, so its boundary $\partial\mathcal{R}$ is a finite union of geodesic arcs. Let us label these arcs $F_1, \ldots, F_r$, oriented anti-clockwise. Each of these arcs F_i is part of an isometric circle $S(g_i)$ of some $g_i \in \Gamma$. In fact the elements $g_1, \ldots, g_r$ form a generating set for the group Γ (see Ford [17]). This will be our canonical generating set. Each arc $F_i \subset \partial\mathcal{R}$ is mapped by g_i to another arc $F_j \subset \partial\mathcal{R}$.

The *net* $\mathcal{N}$ of Γ is the set of translates of $\partial\mathcal{R}$ under Γ. That is,

$$\mathcal{N} = \bigcup_{g\in\Gamma} g(\partial\mathcal{R}).$$

We will make the following three additional assumptions about our non-elementary finitely generated Fuchsian group Γ.

1. Γ contains no parabolic elements.
2. Γ is of the *first kind*. That is, its limit set Λ is the whole of S^1.
3. Γ has the *even corners property*. That is, each isometric circle $S(g_i)$ of a canonical generator g_i lies completely in $\mathcal{N}$.

Example. The simplest example of a group Γ with all the above properties is the one whose fundamental domain $\mathcal{R}$ is the regular $4m$-sided hyperbolic polygon with interior angles $\pi/2m$.

More generally we remark (see Bowen & Series [8]) that the fundamental group of any compact surface of genus two or more has a Fuchsian realisation satisfying the above conditions. We develop this theme (in higher dimensions) in §1·8 and §1·9.

The definition of the orbit $\Gamma(x)$ of the point x under the group Γ extends to the case where x lies on the boundary S^1. As before we define $\Gamma(x) = \{g(x) : g \in \Gamma\}$.

Our aim is to construct a piecewise expanding Markov map $T : S^1 \to S^1$ which is *orbit equivalent* to Γ.

Definition 1·15. *A map $T : S^1 \to S^1$ is said to be* **orbit equivalent** *to the group Γ if for Lebesgue almost all $x \in S^1$ we have*

$$y \in \Gamma(x) \qquad \text{if and only if} \qquad \exists\, p, q \geq 0 \text{ with } T^p(x) = T^q(y).$$

We now construct a Markov partition. Let V be the set of vertices in the net $\mathcal{N}$ which are adjacent to vertices of $\mathcal{R}$ but are not themselves vertices of $\mathcal{R}$. Note that V is finite.

For any $v \in V$, let W_v be the (finite) set of points at infinity (i.e. on S^1) of those geodesics in $\mathcal{N}$ which pass through the vertex v.

Let $W = \bigcup_{v \in V} W_v \subset S^1$, a finite set.

The points in W determine a finite partition $\mathcal{P}$ of S^1 into half open intervals. This will be the Markov partition for our map.

We remark that our assumption that Γ contains no parabolic elements ensures that no vertex of $\mathcal{R}$ lies on S^1 (see Ford [17]). This ensures that our Markov partition $\mathcal{P}$ is *finite* (see Bowen & Series [8]).

If $g_1, \dots, g_r$ is our canonical generating set for Γ, then we have a corresponding collection $A(g_1), \dots, A(g_r)$ of closed arcs, oriented anti-clockwise, of S^1. This collection of arcs is a (measurable) partition of S^1. For convenience let us define $g_{r+1} = g_1$.

Then we can define $T_\Gamma = T : S^1 \to S^1$, called the **Bowen-Series map** corresponding to Γ, by

$$T(x) = g_i(x) \quad \text{if} \quad x \in A(g_i) \setminus A(g_{i+1}) \qquad \text{for } 1 \leq i \leq r.$$

The following two results were proved by Bowen & Series [8].

Proposition 1·40. *(Bowen & Series, [8]) The map $T_\Gamma : S^1 \to S^1$ is orbit equivalent to the group Γ.*

Theorem 1·41. *(Bowen & Series, [8]) $T_\Gamma : S^1 \to S^1$ is a piecewise analytic conformal (eventually) expanding Markov map with respect to the partition $\mathcal{P}$.*

Section 1·8. Higher Dimensional Bowen-Series Maps.

In this section we sketch the generalisation of the construction of Bowen-Series maps to higher dimensions.

Let $\mathbf{D} = \{x \in \mathbf{R}^{n+1} : |x| < 1\}$ be the $(n+1)$-dimensional open unit disc.

Let $S^n = \partial\mathbf{D} = \{x \in \mathbf{R}^{n+1} : |x| = 1\}$ be the n-dimensional sphere.

We can consider $\mathbf{D}$ as the Poincaré disc model of $(n+1)$-dimensional hyperbolic space by giving it the hyperbolic metric

$$ds = \frac{2|dx|}{1-|x|^2} \quad .$$

The boundary S^n represents the sphere at infinity.

In this model, the codimension one geodesic planes of $\mathbf{D}$ are precisely the sectors of those n-dimensional Euclidean hyperspheres which are orthogonal to S^n.

Let $\mathcal{M}(\mathbf{D})$ denote the group of orientation-preserving isometries of $\mathbf{D}$. We call this the Möbius group, and refer to its elements as Möbius transformations.

We remark that $\mathcal{M}(\mathbf{D})$ has a realisation as the matrix group $SO(n,1)$. This is the set of real $(n+1)\times(n+1)$ matrices A satisfying

(i) $A^T\begin{pmatrix}1 & 0\\ 0 & -I_n\end{pmatrix}A = \begin{pmatrix}1 & 0\\ 0 & -I_n\end{pmatrix}$, where A^T denotes the transpose of A.

(ii) $\det A = 1$.

Further details on $\mathcal{M}(\mathbf{D})$ are contained in Ahlfors [1].

By analogy with §1·7, we make the following definition.

Definition 1·16. *An element $g \in \mathcal{M}(\mathbf{D})$ is said to be* **parabolic** *if it has no fixed points in $\mathbf{D}$, and exactly one fixed point on S^n.*

Given $g \in \mathcal{M}(\mathbf{D})$ we define the isometric ball $B(g) = \{x \in \overline{\mathbf{D}} : |g'(x)| > 1\}$, and the isometric sphere $S(g) = \{x \in \overline{\mathbf{D}} : |g'(x)| = 1\}$.

As in the case $n = 1$, provided g is not the identity element then $B(g)$ is in fact a segment of an $(n+1)$-ball, and $S(g)$ is a sector of an n-sphere orthogonal to S^n (for details see Ahlfors [1]).

Each $g \in \mathcal{M}(\mathbb{D})$ is conformal and analytic on its (closed) isometric ball.

For $g \in \mathcal{M}(\mathbb{D})$ we also define $A(g) = \{x \in S^n : |g'(z)| \leq 1\}$. Each $A(g) \subset S^n$ is homeomorphic to a closed n-dimensional disc.

Definition 1·17. *A* **Kleinian group** *is a discrete subgroup of* $\mathcal{M}(\mathbb{D})$.

Note that a Kleinian group is simply a higher dimensional analogue of a Fuchsian group.

The terms *orbit*, *limit set* and *elementary* are defined as in §1·7.

Let Γ be a finitely generated non-elementary Kleinian group.

A fundamental domain $\mathcal{R}$ for Γ is defined as in Definition 1·14, and we make the same canonical choice of $\mathcal{R}$ to be the Ford fundamental domain defined by

$$\mathcal{R} = \mathbb{D} \setminus \bigcup_{g \in \Gamma} \overline{B(g)}.$$

This fundamental domain is a hyperbolic polytope. Its boundary $\partial\mathcal{R}$ is a finite union of n-dimensional faces, $F_1, \ldots, F_r$ say. Each face F_i is part of an isometric sphere $S(g_i)$ of some $g_i \in \Gamma$.

As in the case $n = 1$, the elements $g_1, \ldots, g_r$ form a generating set for the group Γ, and we will use this as our canonical generating set. Each face F_i of $\partial\mathcal{R}$ is mapped by g_i to another face F_j of $\partial\mathcal{R}$.

Let us put an arbitrary ordering $g_1 < g_2 < \ldots < g_r$ on the generating set. This induces an ordering on the faces $F_1 < F_2 < \ldots < F_r$. If we let $A_i = A(g_i)$ then we have an induced ordering $A_1 < A_2 < \ldots < A_r$.

The definition of the net $\mathcal{N}$ of Γ is as in the previous section.

Once more let us impose the following three assumptions on our non-elementary finitely generated Kleinian group Γ.

1. Γ contains no parabolic elements.

2. Γ is of the *first kind*. That is, its limit set Λ is the whole of S^n.

3. Γ has the *even corners property*. That is, each isometric sphere $S(g_i)$ of a canonical generator g_i lies completely in the net $\mathcal{N}$.

We remark (see Rocha [59] for further details) that the fundamental groups of certain compact hyperbolic three-manifolds have a Kleinian realisation satisfying the above hypotheses. (More precisely, their fundamental groups are quasiconformally conjugate to even cornered Kleinian groups). Similar examples can be constructed for manifolds of arbitrarily large dimension. Unlike the well developed two dimensional theory, however, it is not known how prevalent this phenomenon is in higher dimensions.

These assumptions allow us to define a finite Markov partition of S^n as in the last section.

As before, let V be the set of vertices in the net $\mathcal{N}$ which are adjacent to vertices of $\mathcal{R}$ but are not themselves vertices of $\mathcal{R}$. Note that V is a finite set.

For any $v \in V$, let W_v be the set of points at infinity (i.e. on S^n) of those codimension one geodesic planes in $\mathcal{N}$ which pass through the vertex v. Each W_v is a *finite* union of (intersecting) $(n-1)$-spheres embedded in S^n.

Let $W = \bigcup_{v \in V} W_v \subset S^n$. So W is also a *finite* union of (intersecting) $(n-1)$-spheres embedded in S^n. That is, W is an $(n-1)$-dimensional net in S^n. Thus W has measure zero with respect to n-dimensional Lebesgue measure on S^n. So W defines a finite (measurable) partition $\mathcal{P}$ of S^n consisting of the connected components of $S^n \setminus W$.

We use the order on the generators g_i and the 'patches' A_i to define a map $T_\Gamma : S^n \to S^n$ as follows.

Definition 1·18. *The* **Bowen-Series map** $T_\Gamma : S^n \to S^n$ *corresponding to the group* Γ *is given by*

$$T_\Gamma(x) = g_i(x) \quad \text{if } x \in A_i \setminus \bigcup_{j=1}^{i-1} A_j.$$

(Note that this construction works in the case $n = 1$, though in §1·7 we defined T_Γ slightly differently.)

The following results are due to André Rocha [59].

Proposition 1·42. *(Rocha, [59]) The map* $T_\Gamma : S^n \to S^n$ *is orbit equivalent to the group* Γ.

Theorem 1·43. *(Rocha, [59]) $T_\Gamma : S^n \to S^n$ is a piecewise analytic conformal (eventually) expanding Markov map with respect to the partition $\mathcal{P}$.*

Section 1·9. Mostow's Rigidity Theorem.

A version of the celebrated Mostow Rigidity Theorem is the following.

Theorem 1·44. *(Mostow, [42]) Let M_1, M_2 be n-dimensional Riemannian manifolds of constant negative sectional curvature -1, where $n \geq 3$. If M_1, M_2 have the same homotopy type then in fact they are isometric.*

In this section we will give a sketch of Mostow's proof of this result (for further details see Mostow [42] and Thurston [69]), and indicate how Theorem 1·38 can be used to replace part of this proof, for a certain class of manifolds.

First we recall that any n-dimensional manifold M of constant negative curvature -1 has as its universal cover the n-dimensional hyperbolic space D. Here $\mathsf{D} = \{x \in \mathbf{R}^n : |x| < 1\}$ denotes the Poincaré disc model of hyperbolic space, with metric

$$ds = \frac{2|dx|}{1 - |x|^2} \quad .$$

The sphere at infinity is S^{n-1}.

The fundamental group of M is isomorphic to a discrete subgroup Γ of isometries of D (that is, Γ is a Kleinian group), so we can write $M = \mathsf{D}/\Gamma$.

We now assume that Γ is a finitely generated Kleinian group of the first kind, contains no parabolic elements, and satisfies the even corners property. In chapter 2 of Rocha [59] it is shown that a large class of three-dimensional manifolds has this even cornered property. Higher dimensional examples can also be constructed (Jim Anderson, personal communication).

For any $x \in \mathsf{D}$ we let $\Gamma(x)$ denote the equivalence class $\{\gamma(x) : \gamma \in \Gamma\}$.

Let us write $M_i = \mathsf{D}/\Gamma_i$, where Γ_i are Kleinian groups, for $i = 1, 2$.

Suppose $h : M_1 \to M_2$ is a homotopy equivalence. This implies that the fundamental groups $\pi_1(M_1)$ and $\pi_1(M_2)$ are isomorphic. Therefore there exists a group isomorphism $\phi : \Gamma_1 \to \Gamma_2$ between the respective Kleinian groups.

Let $\tilde{h} : \mathbf{D} \to \mathbf{D}$ be a continuous lift of h to the universal cover.

Then $\tilde{h}$ satisfies (see page 103 of Mostow [42]) the *equivariance property*. That is,

$$\tilde{h}(\gamma(x)) = \phi(\gamma)(\tilde{h}(x)) \qquad \text{for all } \gamma \in \Gamma_1, x \in \mathbf{D}.$$

This implies that for all $x \in \mathbf{D}$ we have

$$\tilde{h}(\Gamma_1(x)) = \Gamma_2(\tilde{h}(x)). \tag{1·33}$$

Mostow proved (Theorem 10.1, [42]) that $\tilde{h}$ can be extended continuously to a homeomorphism of the boundary. Let $\tilde{H} : S^{n-1} \to S^{n-1}$ denote this boundary homeomorphism.

By continuity and by equation (1·33) we see that $\tilde{H}$ satisfies

$$\tilde{H}(\Gamma_1(x)) = \Gamma_2(\tilde{H}(x)) \qquad \text{for all } x \in S^{n-1}. \tag{1·34}$$

Using the construction of §1·8 we can find Bowen-Series maps $T_1, T_2 : S^{n-1} \to S^{n-1}$ which are orbit equivalent (see Definition 1·15) to Γ_1, Γ_2. Then equation (1·34) implies that

$$\tilde{H}(T_1(x)) = T_2(\tilde{H}(x)) \qquad \text{for a.e. } x \in S^{n-1}. \tag{1·35}$$

Recall from §1·7 that the maps $T_1, T_2 : S^{n-1} \to S^{n-1}$ are piecewise analytic (in particular C^∞) conformal expanding Markov maps with respect to certain partitions $\mathcal{P}_1, \mathcal{P}_2$ respectively.

The boundary homeomorphism $\tilde{H}$ is a bi-Lipschitz map (see Thurston [69], page 5.39). In particular, Rademacher's Theorem (see Morgan [40], page 27) implies that both $\tilde{H}$ and $\tilde{H}^{-1}$ are differentiable almost everywhere. Moreover, the Lipschitz property means that all partial derivatives of $\tilde{H}$ and $\tilde{H}^{-1}$ are essentially bounded.

Mostow proved (Theorems 9.3 and 10.2, [42]) that in fact $\tilde{H}$ and $\tilde{H}^{-1}$ are both quasiconformal. This means that the images of small spheres have bounded distortion (see page 90, [42] for a definition). In particular, this implies (see Theorem 9.3, [42]) that $\tilde{H}$ and $\tilde{H}^{-1}$ are absolutely continuous on lines. This means that the coordinate functions are absolutely continuous on lines parallel to the coordinate axes (see page 62, [42] for a definition). This absolute continuity condition ensures that

$$\{x \in S^{n-1} : \tilde{H} \text{ is differentiable at } x \text{ and } \tilde{H}^{-1} \text{ is differentiable at } \tilde{H}(x)\}$$

has full Lebesgue measure.

Therefore $\tilde{H}$ is an almost everywhere differentiable conjugacy between the piecewise C^∞ conformal expanding Markov maps T_1 and T_2.

Mostow further shows (pages 99–101, [42]) that $\tilde{H}$ is conformal almost everywhere (that is, there exist almost everywhere defined maps $\alpha : S^{n-1} \to \mathbf{R}$ and $A : S^{n-1} \to SO(n-1)$ such that $D_x\tilde{H} = \alpha(x)A(x)$ for almost every $x \in S^{n-1}$).

He then considers a certain measurable line field associated with $D\tilde{H}$ and uses the ergodicity of the action of Γ_1 on $S^{n-1} \times S^{n-1}$ to show that $\tilde{H}$ is in fact (everywhere) conformal.

However, we can use Theorem 1·38 to give an alternative proof of the conformality of $\tilde{H}$ as follows.

Since $\tilde{H}$ is an almost everywhere differentiable conjugacy between the piecewise C^∞ conformal expanding Markov maps T_1 and T_2, we can apply Theorem 1·38 to deduce that in fact $\tilde{H}$ is a piecewise C^∞ diffeomorphism. That is, $\tilde{H}$ is everywhere C^∞ except possibly on the boundary $\partial\mathcal{P}_1$, and $\tilde{H}^{-1}$ is everywhere C^∞ except possibly on the boundary $\partial\mathcal{P}_2$.

We would like to show that in fact $\tilde{H}$ and $\tilde{H}^{-1}$ are *everywhere* C^∞. Since we have a lot of freedom in our choice of the Bowen-Series maps (recall from §1·8 that we chose an arbitrary ordering of a certain partition of S^{n-1}), we can choose n pairs of Bowen-Series maps $T_1^{(i)}, T_2^{(i)} : S^{n-1} \to S^{n-1}$. We have n pairs of corresponding partitions $\mathcal{P}_1^{(i)}$, $\mathcal{P}_2^{(i)}$ of S^{n-1}, and these may be chosen such that $\cap_{i=1}^n \mathcal{P}_1^{(i)} = \emptyset = \cap_{i=1}^n \mathcal{P}_2^{(i)}$.

By the same argument as above we can show that $\tilde{H}$ is an almost everywhere differentiable conjugacy between each pair of piecewise C^∞ conformal expanding Markov maps $T_1^{(i)}$ and $T_2^{(i)}$. Applying Theorem 1·38 we deduce that $\tilde{H}$ is a piecewise C^∞ diffeomorphism. That is, for each i, $\tilde{H}$ is everywhere C^∞ except possibly on the boundary $\partial\mathcal{P}_1^{(i)}$, and $\tilde{H}^{-1}$ is everywhere C^∞ except possibly on the boundary $\partial\mathcal{P}_2^{(i)}$.

But since $\cap_{i=1}^n \mathcal{P}_1^{(i)} = \emptyset = \cap_{i=1}^n \mathcal{P}_2^{(i)}$, we deduce that $\tilde{H}$ is an *everywhere* C^∞ diffeomorphism. In particular, the derivative $D\tilde{H}$ is continuous.

Now since $\tilde{H}$ is conformal almost everywhere, it is conformal on a dense subset of S^{n-1}. But the matrix group $SO(n-1)$ is a *closed* subset of $M(n-1)$, so the continuity of $D\tilde{H}$ means that in fact $\tilde{H}$ is (everywhere) conformal.

Mostow's Rigidity Theorem follows easily from the conformality of $\tilde{H}$ in the following way. Since $n - 1 \geq 2$, the conformal map $\tilde{H} : S^{n-1} \to S^{n-1}$ is in fact a Möbius transformation (see Lemma 12.2, [42]). But a Möbius transformation of S^{n-1} can be extended to a Möbius transformation of $\mathbf{D}$. Let $\tilde{g} : \mathbf{D} \to \mathbf{D}$ denote the Möbius transformation extending $\tilde{H}$. So $\tilde{g}$ is a hyperbolic isometry. Moreover (see page 103 of [42]), $\tilde{g}$ satisfies

$$\tilde{g}(\Gamma_1(x)) = \Gamma_2(\tilde{g}(x)) \qquad \text{for all } x \in \mathbf{D}.$$

This ensures that $\tilde{g}$ projects down to an isometry $g : M_1 \to M_2$, thus completing the proof of Mostow's Theorem. $\square$

Chapter 2. Cohomological Triviality For Two-Dimensional Subshifts

Section 2·1. Introduction.

Several authors have noted that, in contrast to the situation for $\mathbb{Z}$-actions, the first cohomology of a $\mathbb{Z}^d$-action, for $d > 1$, can be very small. This is one of several *rigidity* properties enjoyed by higher dimensional actions on compact sets X.

If X is a d-dimensional subshift (all terminology will be explained later in the chapter), we can consider the $\mathbb{Z}^d$-action given by the d commuting shift maps. For a group G, we consider cocycles $F : \mathbb{Z}^d \times X \to G$. Typically we impose some kind of regularity condition on F, such as local constancy. If G carries a metric then appropriate regularity assumptions are Hölder continuity or summable variation. Various types of cocycle triviality have been observed as we vary the structure of the subshift X and the group G.

The strongest form of triviality is when F is cohomologous to a homomorphism $\mathbb{Z}^d \to G$ (i.e. a cocycle generated by a function which is constant on X). If G is any group with a doubly invariant metric (such a metric exists, for example, if G is compact, abelian, or discrete) then Klaus Schmidt [63] gives sufficient conditions on X to ensure this strong triviality. These conditions are that X is topologically mixing, and that its Gibbs equivalence relation satisfies a certain specification property (definitions of these terms, and full details of these results, can be found in [63]). By a minor modification of Theorem 3.2 in [63] we can show that if the mixing assumption is replaced by the weaker assumption of topological transitivity, then the same result is true for all *abelian* groups G. In both cases, if the cocycle is Hölder then so is the transfer function, and if the cocycle is of summable variation (a weaker property) then the transfer function is continuous.

If the subshift X itself carries an abelian group structure, then more precise results are possible. For example, Katok & Schmidt [27] show that if such an X is mixing, then all *real*-valued cocycles are strongly trivial. For cocycles taking values in a general group

G, this strong triviality is no longer true. However, Parry and Schmidt conjecture (see Parry [47] and Schmidt [64]) that any cocycle taking values in an *abelian* group G is cohomologous to an *affine* cocycle (i.e. one which is the sum of a homomorphism $\mathbb{Z}^d \to G$ and a homomorphism $X \to G$). In [64] this weaker triviality conjecture is proved for the case $G = S^1$, while [47] gives sufficient conditions for the conjecture to hold if G is *finite*. We remark that much of this algebraic theory holds for any expansive and mixing $\mathbb{Z}^d$-action by automorphisms of a compact abelian group X. Nevertheless, the motivating examples are the subgroups of finite type of the kind first studied by Ledrappier [30].

In a more geometric context, Katok & Spatzier [26] proved that for certain Anosov actions $\mathbb{Z}^d \times X \to X$ on a compact manifold X, all real-valued Hölder cocycles are strongly trivial, with Hölder transfer function.

In this chapter we give a sufficient condition to ensure the strong triviality of cocycles on a $\mathbb{Z}^2$ subshift X. This condition is that the alphabet of the subshift contains some *semi-safe* symbol. Roughly speaking, a symbol is semi-safe if it can be placed next to any other symbol in at least one horizontal direction and at least one vertical direction. Examples of such subshifts are the full shift, the golden mean shift, and the nearest neighbour subshifts considered by Burton & Steif [11]. We do not assume that X carries a group structure, nor that it is of finite type. In contrast to [63] we do not require X to be topologically mixing or even topologically transitive. We give examples of subshifts which are not topologically transitive, do not have transitive Gibbs equivalence relation, yet still have trivial cohomology.

Initially we consider locally constant cocycles taking values in the real numbers (considered as an additive group). These assumptions lead to the consideration of a large system of linear equations. The variables in these equations are the (finite number of) values of the cocycle. We show (Proposition 2·10) that the system of equations contains enough independence to guarantee that *all* variables are expressible in terms of a sufficiently small number of basis variables. This implies cohomological triviality (Theorem 2·12). The method of proof has a geometrical flavour, and is different from the approach of Schmidt [63]. In §2·11 we illustrate our method with a worked example.

In §2·12 we indicate how the results for real-valued locally constant cocycles can be extended to real-valued Hölder cocycles. We first show (Proposition 2·13) that if X is a semi-safe symbol subshift, then the locally constant cocycles are uniformly dense in the space of Hölder cocycles (this fact is not immediate for two-dimensional subshifts). A transitivity assumption allows us to use a result of Livsic [31], and trivial Hölder cohomology (Theorem 2·15) follows by approximation.

For ease of presentation we concentrate mainly on real-valued cocycles, but our methods work for cocycles taking values in a far wider class of groups G. In §2·13 we prove (Theorem 2·23) the triviality of locally constant cocycles with values in any locally (residually finite) group. This class includes all abelian groups, metabelian groups, linear groups, free groups, and many other interesting classes of groups. In particular, certain groups without a doubly invariant metric (for example, general linear groups) are locally (residually finite). Such groups were not considered in [63].

We also discuss briefly the relation between cohomology and topological entropy. While trivial cohomology tends to be associated with the more 'chaotic' or 'mixing' systems, there is no direct link with positive topological entropy. Among our examples of subshifts with trivial cohomology there are both zero entropy and positive entropy systems. Conversely, there are examples (see [63]) of positive entropy subshifts which admit non-trivial Hölder cocycles. Entropy is a quantitative measure of the asymptotic growth rate of the *number* of allowed blocks of a given size. Cohomology, on the other hand, gives a more qualitative idea of how these allowed blocks 'fit together'. The more 'overlap' there is between blocks, the more likely it is that all cocycles will be trivial. This point of view seems in some way linked to the fundamental group of a $\mathbb{Z}^2$-shift defined in Geller & Propp [20]. The extent of this link is not clear, though it is known that under certain conditions (see Corollary 3.6 of [63]), trivial fundamental group implies trivial cohomology.

Section 2·2. Two-Dimensional Subshifts.

Definition 2·1. *For $k \geq 2$ we define our* **alphabet** *to be the finite set $A = \{0, \ldots, k-1\}$. Let $A^{\mathbb{Z}^2}$ denote the set of all maps $x : \mathbb{Z}^2 \to A$. We call this set the* **2-dimensional full shift on** A. *We can think of each $x \in A^{\mathbb{Z}^2}$ as a decoration of the planar lattice by symbols chosen from our alphabet, and we will refer to such x as* **points** *of the set $A^{\mathbb{Z}^2}$.*

We write a typical point $x \in A^{\mathbb{Z}^2}$ as $x = (x_{(m,n)}) = (x_{(m,n)})_{(m,n)\in\mathbb{Z}^2}$, where $x_{(m,n)}$ denotes the value of x at the coordinate (m,n) of the $\mathbb{Z}^2$ lattice.

Note that $A^{\mathbb{Z}^2}$ is compact in the Tychonov product topology, and that the metric δ on $A^{\mathbb{Z}^2}$ given by

$$\delta(x,y) = \begin{cases} 0 & \text{if } x = y \\ 2^{-\min\{|m|+|n| : x_{(m,n)} \neq y_{(m,n)}\}} & \text{otherwise} \end{cases}$$

induces this topology.

Definition 2·2. *The* **horizontal shift** $\sigma : A^{\mathbb{Z}^2} \to A^{\mathbb{Z}^2}$ *and the* **vertical shift** $\tau : A^{\mathbb{Z}^2} \to A^{\mathbb{Z}^2}$ *are defined by:*

$$(\sigma(x))_{(m,n)} = x_{(m+1,n)} \quad , \quad (\tau(x))_{(m,n)} = x_{(m,n+1)} \; .$$

Note that σ and τ are commuting homeomorphisms of $A^{\mathbb{Z}^2}$, so that they define a $\mathbb{Z}^2$ action $\mathbb{Z}^2 \times A^{\mathbb{Z}^2} \to A^{\mathbb{Z}^2}$.

Definition 2·3.

a) A subset $X \subseteq A^{\mathbb{Z}^2}$ is said to be **shift-invariant** *if $\sigma(X) = X$ and $\tau(X) = X$.*

b) A closed, shift-invariant subset $X \subseteq A^{\mathbb{Z}^2}$ is called a 2-dimensional **subshift**.

Note that every subshift is a totally disconnected compact metric space with metric given by the appropriate restriction of δ.

Section 2·3. Rectangles, Blocks, and Cylinder Sets.

Definition 2·4. *Suppose $M, N \in \mathbb{N}$ and $(m_0, n_0) \in \mathbb{Z}^2$. We define the* **rectangle** *$R_{(m_0,n_0)}(M, N)$ (based at (m_0, n_0), of width M, and height N) by*

$$R_{(m_0,n_0)}(M, N) = \{m_0, \ldots, m_0 + M - 1\} \times \{n_0, \ldots, n_0 + N - 1\}.$$

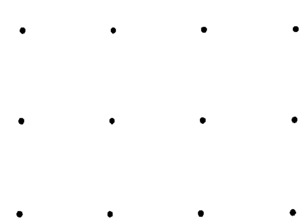
Often the basepoint (m_0, n_0) is unimportant, in which case we simply write $R(M, N)$.

For any $N \in \mathbb{N}$ we define the following squares:

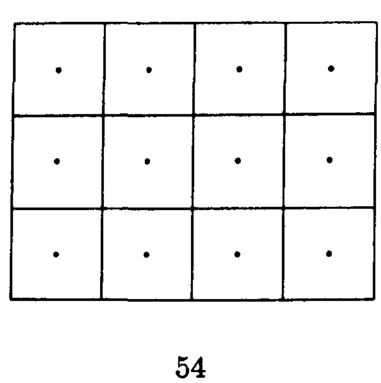
$$S_N = \{1 - N, \ldots, N - 1\} \times \{1 - N, \ldots, N - 1\},$$

$$T_N = \{1 - N, \ldots, N\} \times \{1 - N, \ldots, N\}.$$

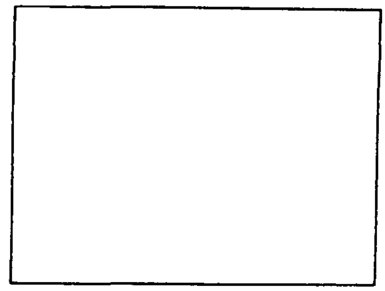
In later sections it will be helpful to draw pictures of rectangles. The rectangle $R(M, N)$ (here $M = 4, N = 3$) is a finite rectangular lattice.



By drawing a unit square around each dot we can identify $R(M, N)$ with a rectangular grid.



When drawing pictures of $R(M,N)$ we will usually suppress the dots and the grid, so we are left with an $M \times N$ rectangle in $\mathbb{R}^2$.



Definition 2·5. *If $F \subseteq \mathbb{Z}^2$ (usually F will be a finite set) then we define $\pi_F : A^{\mathbb{Z}^2} \to A^F$ to be the projection map which restricts each $x \in A^{\mathbb{Z}^2}$ to the set F.*

More generally, and with slight abuse of notation, if $F \subseteq F' \subseteq \mathbb{Z}^2$ then we define $\pi_F : A^{F'} \to A^F$ to be the projection map which restricts each element of $A^{F'}$ to the set F.

Definition 2·6.

(a) Given a finite alphabet A and a subset $F \subseteq \mathbb{Z}^2$, we say that an element of A^F is a **decoration** *of F.*

(b) A decoration B of a rectangle $R \subset \mathbb{Z}^2$ is called a **block**.

(c) Given a subshift $X \subseteq A^{\mathbb{Z}^2}$, we say the block B is **globally allowed** *by X if there exists some $x \in X$ with $\pi_R(x) = B$. If this is the case then we also say that B* **extends to a point** *of X, that x is an* **extension** *of B, and that B* **appears** *in x.*

Let $B \in A^R$ be a block, with corresponding rectangle $R \subset \mathbb{Z}^2$. We can decorate any translated rectangle $R + (i,j)$ in the same way as B, to obtain a block B'. That is, we define $B'_{(m+i,n+j)} = B_{(m,n)}$ for all $(m,n) \in R$.

The shift invariance of X means that B is globally allowed if and only if B' is.

Therefore it is often convenient to think of blocks B without regard to the *position* of the corresponding rectangle R (i.e. only the *size* of R is important). (More precisely we can define the above blocks B, B' to be equivalent, and then quotient out the set of all blocks by this equivalence relation). We will often think of blocks in this way, though for the next definition we *do* take into account the position of the rectangle R.

Definition 2·7. *Given a subshift $X \subseteq A^{\mathbb{Z}^2}$ and a globally allowed block $B \in A^R$, where $R = R_{(m_0,n_0)}(M,N)$, we define the (rectangular)* **cylinder set** $[B] = [B]_{(m_0,n_0)}$ *by*

$$[B] = \{x \in X : \pi_R(x) = B\}.$$

We say that $[B]$ is a cylinder set of **size** R.

If we want to make explicit the block B we will write

$$[B] = \begin{bmatrix} B_{(m_0,n_0+N-1)} & \cdots & B_{(m_0+M-1,n_0+N-1)} \\ \vdots & & \vdots \\ B_{(m_0,n_0)} & \cdots & B_{(m_0+M-1,n_0)} \end{bmatrix}_{(m_0,n_0)} . \tag{2·1}$$

Sometimes the basepoint (m_0, n_0) will be clear from the context (see the spaces V_N in §2·8), in which case we omit the subscript from the right hand square bracket.

We remark that cylinder sets are always non-empty, since we only define them for *globally allowed* blocks. Moreover, they are both open and compact. For a fixed rectangle $R = R_{(m_0,n_0)}$, the family of cylinder sets

$$\{[B]_{(m_0,n_0)} : B \in A^R \text{ is globally allowed }\}$$

determines a finite partition of the subshift X.

Our drawings of rectangles will sometimes be used to represent either blocks (i.e. a decoration of the rectangles) or the corresponding cylinder set. The drawings will be clearly labelled to avoid confusion. Note that whenever we refer to an $M \times N$ rectangle, an $M \times N$ block, or an $M \times N$ cylinder set, the *horizontal* dimension is M, while the *vertical* dimension is N.

Section 2·4. Subshifts of Finite Type.

Definition 2·8. *If $F \subseteq \mathbb{Z}^2$ is finite, and $P \subseteq A^F$, then*

$$X = X_{(F,P)} = \{x \in A^{\mathbb{Z}^2} : \pi_F(\sigma^m \tau^n(x)) \in P \quad \forall (m,n) \in \mathbb{Z}^2\} \tag{2·2}$$

is called the 2*-dimensional* **subshift of finite type** *defined by F and P.*

We call P the set of **locally allowed decorations** *of the finite set F of lattice points.*

We remark that every subshift of finite type is a subshift, and that by the usual recoding argument (see Parry [46] for the one-dimensional case) we may assume that $F = \{0,1\}^2$.

If $X_{F,P}$ is a subshift of finite type, we say a block $B \in A^R$ is **locally allowed** if $\pi_{F' \cap R}(B) \in \pi_{F' \cap R}(P)$ for each translation $F' = F + (i,j)$ which intersects the rectangle R. Every globally allowed block (see Definition 2·6 (c)) is locally allowed, but the converse is not necessarily true. Moreover, and in contrast to the case for one-dimensional subshifts of finite type, in general there exists no finite time algorithm for determining whether a given locally allowed block is globally allowed. Consequently there exists no finite time algorithm for determining whether a given subshift of finite type is the empty set or not. Further discussion of these 'extension' and 'emptiness' problems can be found in Berger [5], Kitchens & Schmidt [28], Robinson [58] and Wang [73].

Certain subshifts of finite type can be defined in terms of two 0-1 square matrices, in the following way.

Definition 2·9. *Let $A = \{0, \ldots, k-1\}$, and suppose M_H, M_V are $k \times k$ zero-one matrices. We define the* **matrix subshift** *$X \subseteq A^{\mathbb{Z}^2}$ by*

$$X = \left\{x \in A^{\mathbb{Z}^2} : M_H(x_{(m,n)}, x_{(m+1,n)}) = 1, \quad M_V(x_{(m,n)}, x_{(m,n+1)}) = 1 \quad \forall (m,n) \in \mathbb{Z}^2\right\}. \tag{2·3}$$

It is easy to check that (2·3) does indeed define a subshift of finite type.

It is well known that every *one-dimensional* subshift of finite type can be specified by a zero-one square matrix. However, it is *not* true that every two dimensional subshift of finite type is a matrix subshift. This is one of the main reasons why the two dimensional theory is much less well understood. Nevertheless, many of the examples we consider will in fact be matrix subshifts.

Section 2·5. Semi-Safe Symbol Subshifts.

Definition 2·10. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a non-empty subshift. A symbol $a \in A$ is called a* **safe symbol** *for X if every globally allowed block can be extended to a point of X by decorating the rest of $\mathbb{Z}^2$ with the symbol a.*

If such a symbol exists, and if every other symbol is globally allowed (to avoid trivialities) then X is called a **safe symbol subshift**.

There are several possible weaker definitions of a safe symbol, where we only require that the symbol extends globally allowed blocks in *two* directions (one horizontal direction and one vertical direction). First we introduce some notation to describe certain semi-infinite regions of $\mathbb{Z}^2$.

Definition 2·11. *Let $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\} \subset \mathbb{Z}^2$ be a rectangle. We define the following regions relative to R.*

$\{(m,n) \in \mathbb{Z}^2 : m < M^- \text{ and } N^- \leq n \leq N^+\}$ *is the* **West strip** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : m > M^+ \text{ and } N^- \leq n \leq N^+\}$ *is the* **East strip** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : n < N^- \text{ and } M^- \leq m \leq M^+\}$ *is the* **South strip** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : n > N^+ \text{ and } M^- \leq m \leq M^+\}$ *is the* **North strip** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : m \leq M^+ \text{ and } n \leq N^+\}$ *is the* **SouthWest quadrant** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : m \geq M^- \text{ and } n \leq N^+\}$ *is the* **SouthEast quadrant** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : m \geq M^- \text{ and } n \geq N^-\}$ *is the* **NorthEast quadrant** *of R.*

$\{(m,n) \in \mathbb{Z}^2 : m \leq M^+ \text{ and } n \geq N^-\}$ *is the* **NorthWest quadrant** *of R.*

Note that R is *not* a subset of any of its strips, but *is* a subset of each of its quadrants.

We have the following technical lemma.

Lemma 2·1. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a subshift, and let $a \in A$. Fix 'Vert' to mean either 'North' or 'South'. Fix 'Horiz' to mean either 'East' or 'West'. The following three conditions are equivalent.*

(a) For any rectangle $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\} \subset \mathbb{Z}^2$, and any globally allowed block $B \in A^R$, we have the following. There exists $x \in X$ which is an extension of B and which decorates the Horiz strip of R with all a's. There exists $y \in X$ which is an extension of B and which decorates the Vert strip of R with all a's.

(b) For any rectangle $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\} \subset \mathbb{Z}^2$, and any globally allowed block $B \in A^R$, there exists $y \in X$ which decorates the rest (i.e. all except R) of the VertHoriz quadrant of R with all a's.

(c) For any rectangle $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\} \subset \mathbb{Z}^2$, and any globally allowed block $B \in A^R$, there exists $x \in X$ which decorates all of $\mathbb{Z}^2$ except the quadrant diagonally opposite *the VertHoriz quadrant of R with all a's.*

Proof. Throughout the proof we will assume, without loss of generality, that the vertical direction is South, and the horizontal direction is West.

$(a) \Rightarrow (b)$ Let B be a globally allowed block with corresponding rectangle $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\}$. Let $x \in X$ be an extension of B which decorates the West strip of R with all a's. For each $M \in \mathbb{N}$, define the rectangle $R_M = \{M^- - M, \ldots, M^+\} \times \{N^-, \ldots, N^+\}$, and the block $B_M = \pi_{R_M}(x)$. Let $y_M \in X$ be an extension of B_M which decorates the South strip of R_M with all a's. So y_M decorates the region

$$C_M = \{(m,n) \in \mathbb{Z}^2 : M^- - M \le m \le M^+, n \le N^+\} \setminus R$$

with all a's. Now the union of all the C_M's is the whole of the SouthWest quadrant except R. So the compactness of X means we can choose a convergent subsequence y_{M_i} whose limit y decorates all of the SouthWest quadrant (except R) with a's. (Note that if X is a subshift of *finite type* then the sequence y_M can be chosen to be eventually constant).

$(b) \Rightarrow (c)$ Let B be a globally allowed block with corresponding rectangle $R = \{M^-, \ldots, M^+\} \times \{N^-, \ldots, N^+\}$. Let $x \in X$ be any extension of B. For each $M \in \mathbb{N}$, define the rectangle $R^M = \{M^-, \ldots, M^+ + M\} \times \{N^-, \ldots, N^+ + M\}$, and the (globally allowed) block $B^M = \pi_{R^M}(x)$. Let y^M be an extension of B^M which decorates the rest (i.e. all except R^M) of the SouthWest quadrant of R^M with all a's. We note that the union (over all M) of such regions is the complement of the NorthEast quadrant of R. The compactness of X means we can choose a convergent subsequence y^{M_i} whose limit y decorates all of $\mathbb{Z}^2$, except the NorthEast quadrant of R, with a's. (Again, if X is a subshift of *finite type* then the sequence y^M can be chosen to be eventually constant).

$(c) \Rightarrow (a)$ This follows immediately, for if R is any rectangle, then its South strip and its West strip are both contained in the complement of its NorthEast quadrant. $\square$

Definition 2·12. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a non-empty subshift. A symbol $a \in A$ is called a* **semi-safe symbol** *for X if it satisfies any of the (equivalent) conditions in Lemma 2·1. We say that it is of* **type VertHoriz**, *where 'Vert'='North' or 'South', and 'Horiz'='East' or 'West'.*

If such a symbol exists, and if every other symbol is globally allowed (to avoid trivialities), we say that X is a **semi-safe symbol subshift**.

Remarks.

1. A safe symbol subshift is also a semi-safe symbol subshift. In fact it is semi-safe of type SouthWest, SouthEast, NorthEast, and NorthWest.

2. Condition (a) in Lemma 2·1 (ostensibly the weakest condition) is the one we use in our proof of cocycle triviality in §2·10, whereas conditions (b) and (c) are easier to visualise. We use condition (b) in the proof of Lemmas 2·2 and 2·3.

Lemma 2·2. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a semi-safe symbol subshift, with semi-safe symbol a. Define the fixed point $\underline{x} \in A^{\mathbb{Z}^2}$ by $\underline{x}_{(m,n)} = a$ for all $(m,n) \in \mathbb{Z}^2$. Then $\underline{x} \in X$.*

Proof. Without loss of generality suppose a is of type SouthWest. Since $X \neq \emptyset$ (see Definition 2·12), we can choose a globally allowed block B, with corresponding rectangle R. By condition (b) of Lemma 2·1 there exists $y \in X$ which decorates the rest (i.e. all except R) of the SouthWest quadrant of R with all a's. Thus $(\sigma^{-1}\tau^{-1})^n(y) \to \underline{x}$ as $n \to \infty$. The shift invariance of X means that $(\sigma^{-1}\tau^{-1})^n(y) \in X$ for all $n \geq 0$. Since X is closed then $\underline{x} \in X$. $\square$

Examples.

1. The full shift $A^{\mathbf{Z}^2}$ on any finite alphabet A is a safe symbol subshift. In fact *every* symbol $a \in A$ is a safe symbol.

2. The matrix subshift $X \subset \{0,1\}^{\mathbf{Z}^2}$ defined by

$$M_H = M_V = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

is a safe symbol subshift, with safe symbol 0. This is known as the *golden mean* subshift.

3. The matrix subshift $X \subset \{0,1\}^{\mathbf{Z}^2}$ defined by

$$M_H = M_V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

is not a safe symbol subshift, but is a semi-safe symbol subshift. The symbol 0 is a semi-safe symbol, of type SouthWest. The symbol 1 is also a semi-safe symbol, of type NorthEast.

4. The matrix subshift $X \subset \{0,1,2\}^{\mathbf{Z}^2}$ defined by

$$M_H = M_V = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is not a safe symbol subshift, but is a semi-safe symbol subshift. The symbol 0 is a semi-safe symbol of type SouthWest. Note that X is not topologically transitive (i.e. no element has a dense σ, τ-orbit), since the symbols 1 and 2 can never appear in the same element of X.

5. Let $X \subset \{0,1\}^{\mathbf{Z}^2}$ be the set of points in which the symbol 1 appears at most once. Then X is a safe symbol subshift, with safe symbol 0, but is not a subshift of finite type.

Section 2·6. Dynamical Properties.

Definition 2·13. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a subshift. Given an element $x \in X$, we define its* **limit set** *$\Lambda(x)$ to be the set of all $y \in X$ for which there exists $(M,N) \in \mathbb{Z}^2$ and a sequence $n_i \to \infty$ such that $(\sigma^M \tau^N)^{n_i}(y) \to x$ as $i \to \infty$.*

We say that x is **attractive** *if $\Lambda(x)$ is dense in X. In this case we also say that the subshift X is* **attractive**.

Remark. The above definition is similar to the definition of the Gibbs equivalence relation in Schmidt [63]. One of the conditions required in [63] to prove triviality of cocycles is that some point $x \in X$ should have a dense Gibbs equivalence class. We remark that this condition is more restrictive than requiring that X is attractive, since the Gibbs equivalence class of a point x is a subset of its limit set $\Lambda(x)$. X being attractive is one of the ingredients of our proof of cohomological triviality and, although we do require additional hypotheses, our method still covers cases not dealt with in [63]. We use the attractive assumption in Lemmas 2·3 and 2·5.

Lemma 2·3. *Suppose $X \subseteq A^{\mathbb{Z}^2}$ is a semi-safe symbol subshift. Then X is attractive.*

Proof. Without loss of generality let us assume that the semi-safe symbol a is of type SouthWest. By Lemma 2·2 we know that $\underline{x} \in X$, where $\underline{x}_{(m,n)} = a$ for all $(m,n) \in \mathbb{Z}^2$. We will show that $\underline{x}$ is attractive.

Suppose $y \in X$. For any $N \geq 1$, define the block $B_N = \pi_{S_N}(y)$. By condition (b) of Definition 2·12 there exists $y_N \in X$ which decorates the rest (i.e. all except S_N) of the SouthWest quadrant of S_N with all a's. Then each $y_N \in \Lambda(\underline{x})$, since $(\sigma^{-1}\tau^{-1})^n(y_N) \to \underline{x}$ as $n \to \infty$.

But $y_N \to y$ as $N \to \infty$. Therefore $\Lambda(\underline{x})$ is dense in X. □

Definition 2·14. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a subshift. The (two dimensional)* **topological entropy** *of X is defined by*

$$h(X) = \lim_{N \to \infty} \frac{1}{|T_N|} \log |\pi_{T_N}(X)| \ ,$$

where $T_N = \{1-N,\dots,N\} \times \{1-N,\dots,N\} \subset \mathbb{Z}^2$.

Remark. Topological entropy measures the asymptotic growth rate of the number of globally allowed blocks of size T_N.

Proposition 2·4. *Every safe symbol subshift of finite type $X \subseteq A^{\mathbb{Z}^2}$ has positive topological entropy.*

Proof. If $X = X_{F,P}$ (see Definition 2·8) then (by a recoding if necessary) we may assume that $F = \{0,1\}^2$.

Let $a \in A$ be a safe symbol. Since every symbol is globally allowed (see Definition 2·10), the following blocks are certainly contained in P, where the asterisk denotes any other symbol:

$$\begin{matrix} a & a \\ a & a \end{matrix}\;,\quad \begin{matrix} a & a \\ a & \star \end{matrix}\;,\quad \begin{matrix} a & a \\ \star & a \end{matrix}\;,\quad \begin{matrix} a & \star \\ a & a \end{matrix}\;,\quad \begin{matrix} \star & a \\ a & a \end{matrix}\;.$$

Given the $2N \times 2N$ square T_N, we want to estimate $|\pi_{T_N}(X)|$, the number of globally allowed decorations of T_N. Let us assume that $N = 3M$ for some $M \in \mathbb{N}$. Then we can divide T_N into $4M^2$ squares of size 3×3, in the obvious way. We can decorate the central coordinate of each 3×3 square arbitrarily, and then decorate the rest of T_N with the safe symbol a, to obtain a block B. By decorating the rest of $\mathbb{Z}^2$ with the safe symbol a, we obtain a point $x \in A^{\mathbb{Z}^2}$. We see that in fact $x \in X$, since each $\pi_F(\sigma^m \tau^n(x))$ is in the form of one of the above five blocks.

So each such decoration gives an element of $\pi_{T_N}(X)$. But there are $|A|^{4M^2}$ such decorations. Therefore

$$\begin{aligned} h(X) &= \lim_{M\to\infty} \frac{1}{|T_{3M}|} \log|\pi_{T_{3M}}(X)| \\ &\geq \lim_{M\to\infty} \frac{1}{(36M^2)} \log|A|^{4M^2} = \frac{1}{9}\log|A| > 0. \quad \square \end{aligned}$$

Remarks.

1. A safe symbol subshift which is *not* of finite type need not have positive entropy (see Example 5 of §2·5).

2. A *semi*-safe symbol subshift of finite type need not have positive entropy (see Example 3 of §2·5).

Section 2·7. Cohomology.

In this section we introduce real-valued cocycles, which we write additively. In §2·13 we consider cocycles taking values in an arbitrary group G.

Definition 2·15. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a subshift. A function $f : X \to \mathbf{R}$ is* **locally constant** *if it only depends on finitely many coordinates (which we call the* **active coordinates***). That is, there exists some finite subset $E = E_f \subset \mathbf{Z}^2$ such that*

$$\pi_E(x) = \pi_E(y) \quad \Rightarrow \quad f(x) = f(y). \tag{2·4}$$

Remark. We make no requirement that E be the *smallest* set satisfying (2·4), so any such E is just *some* set of active coordinates.

We note that all locally constant functions are Hölder continuous (see §2·12).

Definition 2·16. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a subshift. Suppose $f, g : X \to \mathbf{R}$ are (locally constant) functions. The pair of functions (f, g) is said to be a (locally constant)* **cocycle on X** *if the following relation is satisfied:*

$$f\tau - f = g\sigma - g. \tag{2·5}$$

The following definition of cocycle triviality corresponds to the notion of strong triviality discussed in §2·1.

Definition 2·17. *A (locally constant) cocycle (f, g) on a subshift X is said to be* **trivial** *if there exist constants $c_f, c_g \in \mathbf{R}$ and some function $h : X \to \mathbf{R}$ such that*

$$f = c_f + h\sigma - h \quad \textit{and} \quad g = c_g + h\tau - h. \tag{2·6}$$

Such an h is called a **transfer function**.

Definition 2·18. *If every locally constant cocycle on a subshift X is trivial then we say that X has* **trivial locally constant cohomology**, *or that it is* **cohomologically trivial** *(in the locally constant class).*

We remark that our notation follows that adopted by Parry [48], but is equivalent to the more standard notation employed in, for example, Schmidt [63]. The standard definition of a cocycle is a map $F : \mathbf{Z}^2 \times X \to \mathbf{R}$ satisfying

$$F(m+m', n+n', x) = F(m,n,x) + F(m', n', \sigma^m \tau^n(x))$$

for all $(m,n), (m',n') \in \mathbf{Z}^2$ and $x \in X$. Note that such a map F is generated by a pair of functions (f,g) satisfying (2·5), simply by defining $F(1,0,x) = f(x)$ and $F(0,1,x) = g(x)$. We say that two cocycles F, F' are *cohomologous* if there exists some (transfer) function $h : X \to \mathbf{R}$ such that

$$F(m,n,x) = F'(m,n,x) + h(\sigma^m \tau^n(x)) - h(x)$$

for all $(m,n) \in \mathbf{Z}^2$ and $x \in X$. A particularly simple kind of cocycle is one which is independent of the X variable. We call this a *homomorphism*, and note that it is generated by a pair of constant functions (c_f, c_g). Thus our definition of a trivial cocycle corresponds to one which is cohomologous to a homomorphism.

If X is a semi-safe symbol subshift, then it contains a fixed point, by Lemma 2·2. If (f,g) is a trivial cocycle on X then the constants c_f, c_g are just the respective values of f and g at this fixed point. In particular, c_f and c_g are unique. In fact, if μ is a Borel probability measure on X, invariant under both σ and τ, then $c_f = \int f d\mu$ and $c_g = \int g d\mu$. The fact that *all* probability measures invariant under both σ and τ integrate to the same value is a serious restriction on the set of such measures. Thus cocycle triviality (for $\mathbf{Z}^2$ actions) is linked to the scarcity of invariant measures (for $\mathbf{Z}^2$ actions). This latter phenomenon was first studied by H. Furstenberg [18].

The transfer function h for a trivial cocycle is *not* unique, since adding a constant to h also gives a transfer function. However, in most interesting cases this is the only way of obtaining new transfer functions, as the following lemma demonstrates.

Lemma 2·5. *Suppose $X \subseteq A^{\mathbb{Z}^2}$ is an attractive subshift. Suppose (f, g) is a trivial locally constant cocycle on X with transfer function h. Then h is unique up to an additive constant.*

Proof. Since f and g are locally constant then so is h, by equation (2·6). In particular h is continuous. Let h' be some other transfer function for (f, g). Again we have that h' is locally constant, and therefore continuous.

From (2·6) we have

$$f = c_f + h\sigma - h \quad (1) \qquad g = c_g + h\tau - h \quad (2)$$
$$f = c_f + h'\sigma - h' \quad (1') \qquad g = c_g + h'\tau - h' \quad (2').$$

Subtracting $(1')$ from (1), and $(2')$ from (2), we see that the continuous function $h - h'$ is invariant under both σ and τ.

Suppose $x \in X$ is attractive, and $z \in \Lambda(x)$. So there exists $(M, N) \in \mathbb{Z}^2$ and a sequence $n_i \to \infty$ such that $(\sigma^M \tau^N)^{n_i}(z) \to x$ as $i \to \infty$. Since $h - h'$ is continuous, and invariant under both σ and τ, we deduce that $(h - h')(z) = (h - h')(x)$.

Since $\Lambda(x)$ is dense in X, the continuity of $h - h'$ implies that $(h - h')(y) = (h - h')(x)$ for all $y \in X$. That is, $h - h'$ is a constant function. □

Remarks.

1. By a similar argument we can show that if X is topologically transitive then the transfer function of a trivial cocycle is unique up to an additive constant.

2. In Lemma 2·18 we prove a generalisation of Lemma 2·5, for locally constant cocycles taking values in an arbitrary group. The proof there is slightly more involved.

Corollary 2·6. *Suppose $X \subseteq A^{\mathbb{Z}^2}$ is a semi-safe symbol subshift. Suppose (f, g) is a trivial locally constant cocycle on X with transfer function h. Then h is unique up to an additive constant.*

Proof. Immediate from Lemmas 2·3 and 2·5. □

Section 2·8. Cocycles of Degree N.

Given a locally constant cocycle (f, g) on a subshift X, we would like to specify the active coordinates of both f and g. Since the active coordinates are not chosen to be minimal (see the remark after Definition 2·15), we have a lot of freedom in this. In particular, it will be convenient to assume that the active coordinates for f lie in some rectangle F_N of size $2N \times (2N - 1)$, and that the active coordinates for g lie in some rectangle G_N of size $(2N - 1) \times 2N$, and that the bottom left corners of these rectangles coincide.

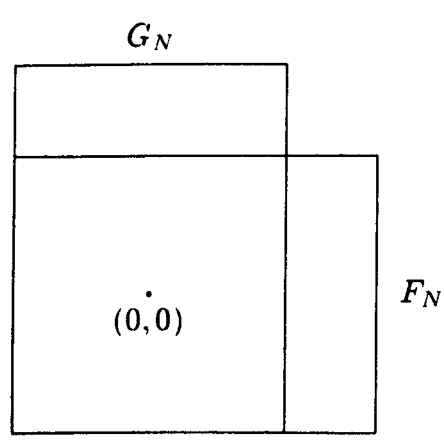
In Definition 2·4 we introduced the sequence of squares S_N, which are symmetric about both axes, and whose union is all of $\mathbb{Z}^2$. We now define F_N and G_N to be the rectangles obtained by adding either a row or a column to S_N. Recall that the square T_N (see Definition 2·4) is obtained by adding a row *and* a column *and* a corner to S_N.

Definition 2·19. *For $N \in \mathbb{N}$ let us define*

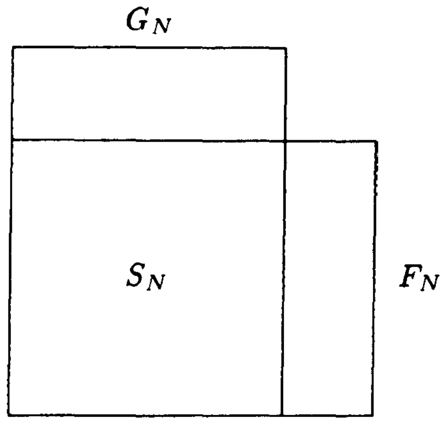
$$F_N = \{1 - N, \ldots, N\} \times \{1 - N, \ldots, N - 1\},$$

$$G_N = \{1 - N, \ldots, N - 1\} \times \{1 - N, \ldots, N\}.$$

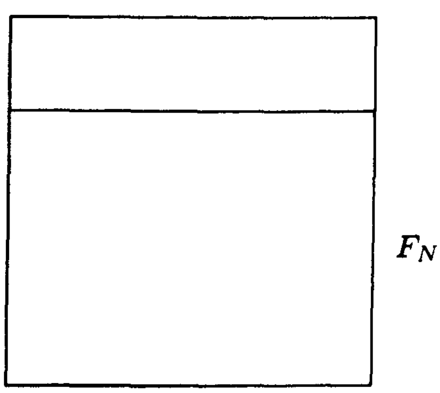
If (f, g) is a locally constant cocycle then we can choose $N \in \mathbb{N}$ such that the active coordinates for f lie in F_N and the active coordinates for g lie in G_N.



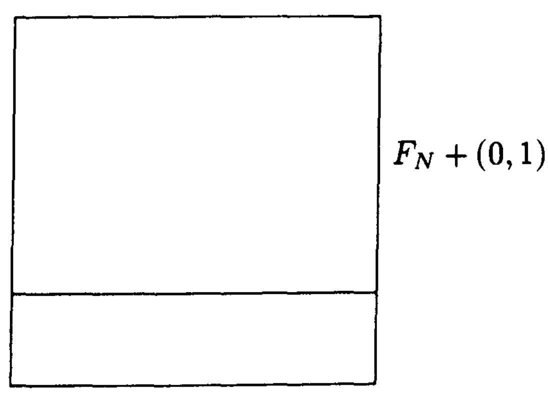
If (f, g) is a *trivial* locally constant cocycle (see Definition 2·17) then the associated transfer function h will also be locally constant. In fact (see (2·6)) its active coordinates will lie in the square S_N.



This is one of the reasons why it is convenient to work with the rectangles F_N and G_N as the active coordinates of f and g. Another reason is that the active coordinates for the four functions f, g, $f\tau$, $g\sigma$ (i.e. the four terms in the cocycle equation (2·5)) all lie in the square T_N. The active coordinates of $f\tau$ lie in $F_N + (0, 1)$, while the active coordinates of $g\sigma$ lie in $G_N + (1, 0)$.

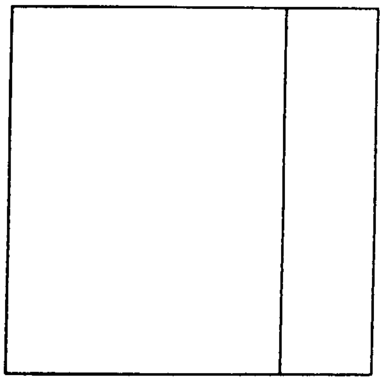


The square T_N



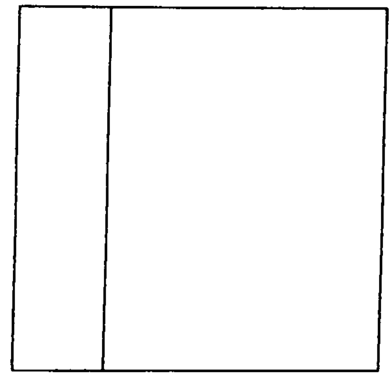
The square T_N

G_N



The square T_N

$G_N + (1,0)$



The square T_N

The *essential* point is simply that both sequences of rectangles get arbitrarily large, so that their union over all N is the whole of $\mathbf{Z}^2$.

Definition 2·20. *A locally constant cocycle (f,g) on a subshift X is said to be of* **degree** N *if the active coordinates of f lie in F_N and the active coordinates of g lie in G_N.*

Note that every locally constant cocycle is of degree N for all sufficiently large N.

Definition 2·21. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a subshift. Let $V_N = V_N(X)$ denote the set of cocycles of degree N. Let $V'_N = V'_N(X)$ denote the subspace of V_N consisting of trivial cocycles.*

Note that V_N and V'_N are both finite dimensional vector spaces over $\mathbf{R}$. In §2·13, where we look at cocycles with values in more general groups, the analogues of V_N and V'_N will have less algebraic structure.

Section 2·9. The System of Linear Cocycle Equations.

We know (see the remarks after Definition 2·7) that for a fixed rectangle $R = R_{(m_0,n_0)}$, the family of cylinder sets $\{[B] : B \in A^R \text{ is globally allowed }\}$ determines a finite partition of the subshift X. It follows that if some function $\phi : X \to \mathbf{R}$ is locally constant, with active coordinates lying in R, then ϕ is completely determined by its values on the (finite number of) cylinder sets of size R. We want to study the dimension of the space V_N of cocycles (f, g) of degree N. That is, we will consider the degrees of freedom we have in assigning values to f and g on their respective cylinder sets. The cocycle equation (2·5) will impose restrictions on this freedom in the form of a system of linear equations (see Lemma 2·7). For this reason we make the following definition.

Definition 2·22. *Suppose $\phi : X \to \mathbf{R}$ is locally constant, with active coordinates lying in $R = R_{(m_0,n_0)}(M, N)$. The value of ϕ on the cylinder set $[B]$ of size R is called the* **variable** *(or ϕ-variable) corresponding to $[B]$, and will be denoted by*

$$\phi([B]) = \{B\}_\phi = \left\{ \begin{matrix} B_{(m_0,n_0+N-1)} & \cdots & B_{(m_0+M-1,n_0+N-1)} \\ \vdots & & \vdots \\ B_{(m_0,n_0)} & \cdots & B_{(m_0+M-1,n_0)} \end{matrix} \right\}_\phi . \tag{2·7}$$

This notation will only ever be used in the context of a cocycle (f, g) and transfer function h. Since the *size* of the active coordinates rectangle is *different* for each of these functions ($2N \times (2N-1)$ for f, $(2N-1) \times 2N$ for g, $(2N-1) \times (2N-1)$ for h), we will sometimes omit the subscript from the right hand bracket in (2·7) without causing any ambiguity.

Note the difference between the notation in (2·1) and in (2·7). We will always use *square* brackets to denote the cylinder set itself, and *curly* brackets to denote the value of a function on the cylinder set.

The following elementary lemma is the key to our methods in §2·10.

Lemma 2·7. *Suppose $(f, g) \in V_N(X)$, where $X \subseteq A^{\mathbf{Z}^2}$ is a subshift. The $|\pi_{F_N}(X)|$ f-variables and $|\pi_{G_N}(X)|$ g-variables satisfy a system of $|\pi_{T_N}(X)|$ linear equations. In* *each equation there are two f-variables and two g-variables.*

Proof. Since (f,g) is of degree N, each of the four functions in the cocycle equation has active coordinates in the square T_N (see the discussion after Definition 2·19). Let the block

$$C = \begin{matrix} c_{1-N,N} & \cdots & \cdots & c_{N,N} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ c_{1-N,1-N} & \cdots & \cdots & c_{N,1-N} \end{matrix}$$

be a globally allowed decoration of T_N, and let $[C]$ be the corresponding cylinder set.

We now consider the cocycle equation $f\tau - f = g\sigma - g$, restricted to the set $[C]$. This simply gives

$$f\tau([C]) - f([C]) = g\sigma([C]) - g([C]).$$

More explicitly we can write this equation as

$$\left\{ \begin{matrix} c_{1-N,N} & \cdots & \cdots & c_{N,N} \\ \vdots & & & \vdots \\ c_{1-N,2-N} & \cdots & \cdots & c_{N,2-N} \end{matrix} \right\}_f - \left\{ \begin{matrix} c_{1-N,N-1} & \cdots & \cdots & c_{N,N-1} \\ \vdots & & & \vdots \\ c_{1-N,1-N} & \cdots & \cdots & c_{N,1-N} \end{matrix} \right\}_f$$

$$= \left\{ \begin{matrix} c_{2-N,N} & \cdots & c_{N,N} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ c_{2-N,1-N} & \cdots & c_{N,1-N} \end{matrix} \right\}_g - \left\{ \begin{matrix} c_{1-N,N} & \cdots & c_{N-1,N} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ c_{1-N,1-N} & \cdots & c_{N-1,1-N} \end{matrix} \right\}_g .$$

This is a linear equation in two f-variables and two g-variables. $\square$

Example. Let $X = \{0,1\}^{\mathbf{Z}^2}$ be the full shift on two symbols, and suppose $(f,g) \in V_2(X)$. So the cocycle equation gives a system of $2^{16} = 65536$ equations in $2^{12} + 2^{12} = 8192$ variables. For example the block

$$C = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{matrix}$$

gives rise to the equation

$$\begin{Bmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{Bmatrix} - \begin{Bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{Bmatrix} = \begin{Bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{Bmatrix} - \begin{Bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{Bmatrix}.$$

The important point is that if we already *know* 3 of the variables in some linear cocycle equation, then the equation allows us to deduce the fourth variable as well. We use this idea in the proof of Proposition 2·10. In §2·13 we use this same idea, but with a slightly modified argument, to consider locally constant cocycles taking values in groups other than the reals.

In §2·10 we prove that if X is a semi-safe symbol subshift then the spaces $V_N(X)$ and $V'_N(X)$ coincide, by showing they have the same dimension. The next lemma and corollary give an expression for $\dim V'_N(X)$.

Lemma 2·8. *Let $X \subseteq A^{\mathbf{Z}^2}$ be an attractive subshift. Then $\dim V'_N(X) = |\pi_{S_N}(X)|+1$ (where $|\pi_{S_N}(X)|$ is the number of (non-empty) cylinder sets of size S_N).*

Proof. We have 2 degrees of freedom in choosing the constants c_f and c_g (see Definition 2·17).

The active coordinates of the transfer function h lie in the square S_N. There are $|\pi_{S_N}(X)|$ cylinder sets of size S_N, so we have $|\pi_{S_N}(X)|$ degrees of freedom in choosing the variables of h (i.e. the values it takes on the cylinder sets).

However, since X is attractive, we know (by Lemma 2·5) that h is only unique up to an additive constant (i.e. $h + c$ will also be a transfer function, for any $c \in \mathbf{R}$).

So in total there are

$$2 + |\pi_{S_N}(X)| - 1$$

degrees of freedom in defining a trivial cocycle. □

Corollary 2·9. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a semi-safe symbol subshift. Then $\dim V'_N(X) = |\pi_{S_N}(X)| + 1$.*

Proof. Immediate from Lemmas 2·3 and 2·8. □

Section 2·10. Cocycle Triviality For Semi-Safe Symbol Subshifts.

The following Proposition 2·10 is the main result of this chapter. The method of proof is illustrated by the worked example in §2·11. In Proposition 2·11 we indicate the amendments necessary to prove the same result for subshifts with larger alphabets. We remark that the method used in the proof of Proposition 2·10 is quite general. With suitable modifications of notation and terminology, it is valid for locally constant cocycles taking values in *any* group (see Proposition 2·20).

Proposition 2·10. *Let $X \subseteq \{0,1\}^{\mathbb{Z}^2}$ be a semi-safe symbol subshift. Then $\dim V'_N(X) = \dim V_N(X)$ for each $N \in \mathbb{N}$.*

Proof. Without loss of generality let us assume that 0 is the semi-safe symbol, and that it is of type SouthWest.

By Corollary 2·9 we know that $\dim V'_N = |\pi_{S_N}(X)| + 1$.

Suppose $(f,g) \in V_N$. We will fix $|\pi_{S_N}(X)|$ of the f-variables and one of the g-variables (see Definition 2·22), and we will refer to these as the **basis variables**. We claim that the fixing of these $|\pi_{S_N}(X)| + 1$ variables completely determines the cocycle (f,g). We will use our basis variables, together with the system of linear equations (see Lemma 2·7) derived from the cocycle equation (2·5), to determine *all other* variables of f and g. This will imply that we have $|\pi_{S_N}(X)| + 1$ degrees of freedom in our choice of the cocycle (f,g). In other words, $\dim V_N = |\pi_{S_N}(X)| + 1$, thus establishing the proposition.

We choose our basis variables to be the following:

All f-variables of the form

$$\left\{\begin{matrix} 0 & \star & \dots & \star \\ \vdots & \vdots & & \vdots \\ 0 & \star & \dots & \star \end{matrix}\right\}_f$$

(i.e. the values of f on all those cylinder sets of size F_N whose left-hand column is decorated by the semi-safe symbol 0).

There are precisely $|\pi_{S_N}(X)|$ such variables. This is because (by shift invariance) there are $|\pi_{S_N}(X)|$ globally allowed ways of decorating the translated square $S_N + (1,0)$ (i.e. of filling in the asterisks in the above diagram). Since 0 is a SouthWest safe symbol, we can decorate the left hand column of F_N with all 0's, and the resulting block will correspond to a non-empty cylinder set.

The final basis variable is the g-variable consisting solely of 0's:

$$\begin{Bmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{Bmatrix}_g .$$

Note that this all 0's decoration is indeed a variable (i.e. does correspond to a non-empty cylinder set), since 0 is a semi-safe symbol (see Lemma 2·2).

We remark that this choice of basis is canonical, though clearly there are other possible bases. In systems without a semi-safe symbol, a non-trivial problem is to find a canonical way of choosing the basis variables.

Our method of proof is as follows. By Lemma 2·7 we know that each globally allowed block C of size T_N gives a linear equation in four variables. Starting with our basis variables (the 'known' variables) we choose an appropriate block C (i.e. one for which exactly 3 of the variables $f([C])$, $f\tau([C])$, $g([C])$, $g\sigma([C])$ are 'known'). Using the linear cocycle equation on $[C]$ we obtain an expression for the previously unknown variable in terms of the known variables. We now include this variable in the set of known variables.

We repeat the process. As the number of known variables increases, it becomes easier to find appropriate blocks C.

The above discussion is valid for *any* finite alphabet A. From now on, however, we use the fact that $A = \{0, 1\}$. In the proof of Proposition 2·11 we indicate the minor modifications necessary for larger alphabets.

For $A = \{0, 1\}$ we claim that the following statement $P(r)$ is true for all $r \geq 0$.

All variables whose decorations contain r 1's can be expressed in terms of those basis variables whose decorations contain $\leq r$ 1's.

We note that if $P(r)$ is true for all $r \geq 0$ then it follows that all variables are expressible in terms of the $|\pi_{S_N}(X)| + 1$ basis variables (since every variable contains r 1's, for some

$0 \leq r \leq 2N(2N-1)$), and the proposition will have been proved.

We will prove the statement $P(r)$ by induction on r.

Clearly $P(0)$ is true, since the f-variable consisting of all 0's and the g-variable consisting of all 0's are both basis variables.

Let our inductive hypothesis be that $P(j)$ is true for $j = 0, 1, \ldots, r-1$. We will show this implies that $P(r)$ is true.

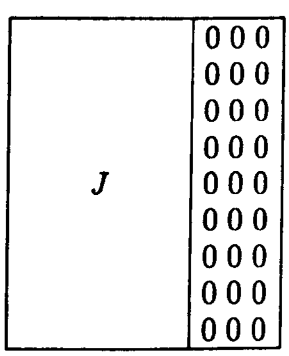
Suppose we know all the *basis* variables with $\leq r$ 1's. By the inductive hypothesis this implies we know *all* variables with j 1's, for $j = 0, 1, \ldots, r-1$. So in total the known variables are:

1. All variables with strictly less than r 1's.
2. All basis variables with r 1's.

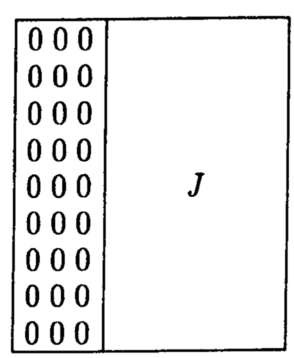
We will try to deduce all variables with r 1's. First we will deduce those g-variables with r 1's, then we will deduce those f-variables with r 1's.

Let B be the $(2N-1) \times 2N$ block corresponding to an arbitrary g-variable $\{B\}$ with r 1's. Let J be the *largest* rectangular sub-block of B which has at least one 1 in its right-hand column (so possibly $J = B$). Note that J also has r 1's.

Consider the block J on its own, and then add columns of zeros to its left (possible since 0 is a semi-safe symbol of type SouthWest) until we have a block of size $(2N-1) \times 2N$. Call this block B_0. Note that B_0 has r 1's as well.

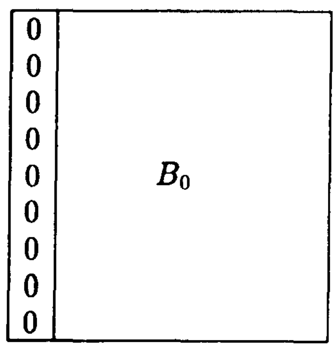


The Block B

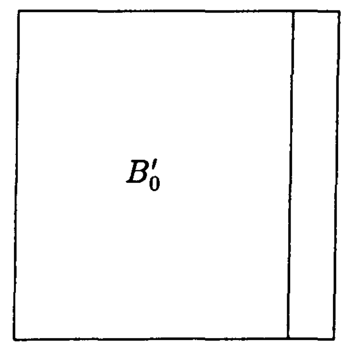


The Block B_0

Now add one further column of zeros to the left of B_0 to make a $2N \times 2N$ square block C_0. Let B'_0 denote the $(2N-1) \times 2N$ block obtained by removing the *right*-hand column of C_0. Since J has at least one 1 in its right-hand column then B'_0 contains strictly less than r 1's.



The Block C_0



The Block C_0

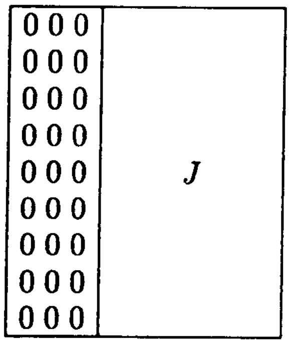
Now consider the cocycle equation on the cylinder set $[C_0]$.

Both f-variables have their left-hand column full of 0's, thus they are basis variables. Moreover, they both have $\leq r$ 1's. Therefore they are known variables.

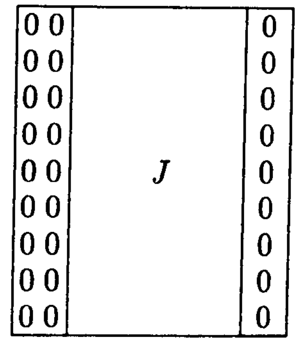
The g-variable $\{B'_0\}$ contains *strictly* less than r 1's, and therefore is a known variable (by the inductive hypothesis).

Thus the only unknown variable is the g-variable $\{B_0\}$. The cocycle equation therefore allows us to deduce $\{B_0\}$, which we now consider a known variable.

If $B_0 = B$ then we are done. Otherwise we can remove a column of zeros from the left of B_0, and add a column of zeros to the right, to obtain a $(2N-1) \times 2N$ block B_1.

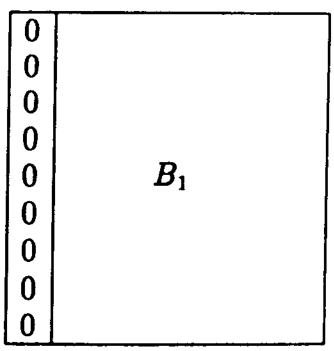


The Block B_0

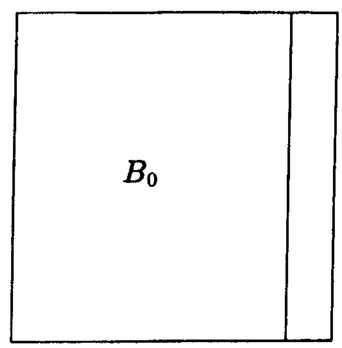


The Block B_1

Let C_1 denote the $2N \times 2N$ square block obtained by adding a column of 0's to the left of B_1. Note that the $(2N-1)\times 2N$ block obtained by removing the right-hand column of C_1 is precisely B_0.



The Block C_1



The Block C_1

Now consider the cocycle equation on the cylinder set $[C_1]$. Both f-variables have their left-hand column full of 0's, thus they are basis variables. Moreover, they both have $\leq r$ 1's. Therefore they are known variables. The g-variable $\{B_0\}$ is also now a known variable. The cocycle equation therefore allows us to deduce the previously unknown g-variable $\{B_1\}$.

In the same way we can continue to define $(2N-1) \times 2N$ blocks $B_2, B_3, B_4 \ldots$ and $2N \times 2N$ square blocks $C_2, C_3, C_4 \ldots$. The cocycle equation on $[C_i]$ always allows us to deduce the variable $\{B_i\}$, since $\{B_{i-1}\}$ is known, and the two f-variables are also known. Eventually some B_i is equal to our original block B, so we have deduced the g-variable $\{B\}$.

Since $\{B\}$ was an arbitrary g-variable with r 1's, we can consider all such variables to now be known.

Now let $\{D\}$ be an unknown f-variable with r 1's, and with corresponding $2N \times (2N-1)$ block D (so the left-hand column of D does *not* consist solely of zeros). Let K be the *largest* rectangular sub-block of D which has at least one 1 in its top row (so possibly $K = D$). Note that K also has r 1's.

Consider the block K on its own, and then add rows of zeros to its bottom until we have a block of size $2N \times (2N-1)$. Call this block D_0. Note that D_0 has r 1's as well.

Now add one further row of zeros to the bottom of D_0 to make a $2N \times 2N$ square block E_0. Let D_0' denote the $2N \times (2N-1)$ block obtained by removing the top row of E_0. Since K has at least one 1 in its top row then D_0' contains strictly less than r 1's.

Now consider the cocycle equation on the cylinder set $[E_0]$. Both g-variables are known, by our previous discussion in this proof. The f-variable $\{D_0'\}$ is also known, since it contains strictly less than r 1's. Thus the cocycle equation allows us to deduce the previously unknown variable $\{D_0\}$, which we now consider a known variable.

If $D_0 = D$ then we are done. Otherwise we can remove a row of zeros from the bottom of D_0 and add a row of zeros to the top, to obtain a $2N \times (2N-1)$ block D_1. Let E_1 denote the $2N \times 2N$ square block obtained by adding a row of 0's to the bottom of D_1. Note that the $2N \times (2N-1)$ block obtained by removing the top row of E_1 is precisely D_0.

Now consider the cocycle equation on the cylinder set $[E_1]$. Both g-variables are known, and the f-variable $\{D_0\}$ is also now a known variable. The cocycle equation therefore allows us to deduce the previously unknown f-variable $\{D_1\}$.

In the same way we can continue to define $2N \times (2N-1)$ blocks $D_2, D_3, D_4 \ldots$ and $2N \times 2N$ square blocks $E_2, E_3, E_4 \ldots$. The cocycle equation on $[E_i]$ always allows us to deduce the variable $\{D_i\}$, since $\{D_{i-1}\}$ is known, and the two g-variables are also known. Eventually some D_i is equal to our original block D, so we have deduced the f-variable $\{D\}$.

Since $\{D\}$ was an arbitrary (non-basis) f-variable with r 1's, we can consider all such variables to now be known.

Therefore all variables (both f-variables and g-variables) with r 1's can be deduced from those basis variables with $\leq r$ 1's. This completes the induction, and the proposition is proved. $\square$

We now generalise Proposition 2·10 to semi-safe symbol subshifts with larger alphabets.

Proposition 2·11. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a semi-safe symbol subshift. Then* $\dim V'_N(X) = \dim V_N(X)$ *for each* $N \in \mathbf{N}$.

Proof. Suppose $A = \{0, \ldots, k-1\}$. As in Proposition 2·10, let us assume that the symbol 0 is the semi-safe symbol, and that it is of type SouthWest.

The case $k = 2$ was dealt with in Proposition 2·10. The method of proof for the general case is almost the same. The only difference is that we must use induction more carefully on the number of symbols of each type appearing in a variable.

As before, we deduce all variables consisting of 0's and 1's from those *basis* variables consisting of 0's and 1's. The method is to use induction on the number of 1's in each variable.

Now we introduce the symbol 2. That is, we consider variables containing only the symbols 0, 1 and 2. Using induction on the number of 2's appearing in such variables, we eventually deduce all variables containing only 0, 1 and 2.

We continue in this manner, introducing one new symbol at a time, until eventually we are able to deduce all variables, thus completing the proof.

(We remark that this proof amounts to putting a lexicographic ordering on the set of k-tuples $\underline{r} = (r_0, r_1, \ldots, r_{k-1})$, where r_i is the number of times the symbol i occurs in a variable. We formulate a statement $P(\underline{r})$ analogous to the statement $P(r)$ in Proposition

2·10, then prove it using induction on $\underline{r}$.) □

Theorem 2·12. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a semi-safe symbol subshift. Then every locally constant cocycle (f, g) on X is a trivial cocycle.*

Proof. By Proposition 2·11 we know that $\dim V'_N = \dim V_N$ for each $N \in \mathbb{N}$. Since V'_N is a subspace of V_N, this implies that $V'_N = V_N$. That is, every cocycle of degree N is trivial. But every locally constant cocycle is of degree N, for some $N \in \mathbb{N}$. Thus every locally constant cocycle is trivial. □

Examples. All the examples in §2·5 are semi-safe symbol subshifts, and therefore have trivial cohomology by Theorem 2·12. Note that examples 3 and 4 are not covered by the techniques of Schmidt [63], since in neither case is there an element with dense Gibbs equivalence class. In fact example 4 is not even topologically transitive.

Section 2·11. The Full Shift - Worked Example.

Let $X = \{0,1\}^{\mathbb{Z}^2}$ be the full shift on two symbols. Both 0 and 1 are safe symbols, but for this example we will use 0 as our safe symbol. In particular, we will consider 0 as a semi-safe symbol of type SouthWest. Let (f, g) be a locally constant cocycle on X, where the active coordinates of f and g lie in the rectangles $\{0,1,2\} \times \{0,1\}$ and $\{0,1\} \times \{0,1,2\}$ respectively. Note that these rectangles are not of the form F_N or G_N (see Definition 2·19), but are of a convenient size to illustrate the proof of Proposition 2·10, without introducing unnecessary computation. (F_1 and G_1 are too small to show why we need the rectangles J and K, whereas if we use F_2 and G_2 then we already have $2^9 = 512$ basis variables).

If (f, g) were a trivial cocycle, its transfer function h would have active coordinates in the square $\{0,1\}^2 \subset \mathbb{Z}^2$. Since there are 2^4 globally allowed decorations of $\{0,1\}^2$, the space of such transfer functions is $2^4 + 1 = 17$ dimensional (cf. Lemma 2·8 and Corollary 2·9).

Therefore to prove the cocycle triviality of X (for cocycles with active coordinates in the above rectangles), we will fix 17 basis variables, and deduce all other f- and g-variables from these.

The variables we fix are as follows. We group them according to the number of times the symbol 1 occurs in them. Note that one of the basis variables is a g-variable, while the other 16 are f-variables. Each f-variable has a left-hand column of zeros, and the right-hand 2×2 square is decorated in one of the 16 allowed ways.

$$\begin{Bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{Bmatrix},$$

$$\begin{Bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{Bmatrix},$$

$$\begin{Bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{Bmatrix}, \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{Bmatrix}, \begin{Bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{Bmatrix},$$

$$\begin{Bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{Bmatrix}, \quad \begin{Bmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \end{Bmatrix},$$

$$\begin{Bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \end{Bmatrix}.$$

As in Proposition 2·10, we start by deducing all g-variables with one 1. To do this we only need use those basis variables with one 1 or no 1's.

For example, suppose we want to deduce the variable

$$\{B\} = \begin{Bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{Bmatrix}.$$

We first define the blocks

$$B_0 = \begin{matrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{matrix} \quad \text{and} \quad C_0 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}.$$

The cocycle equation on the cylinder set $[C_0]$ is

$$\left\{\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{matrix}\right\} - \left\{\begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{matrix}\right\} = \left\{\begin{matrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{matrix}\right\} - \left\{\begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix}\right\}.$$

Since three of these variables are in the basis, we may deduce the unknown variable $\{B_0\}$.

Now define the block

$$C_1 = \begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}.$$

The cocycle equation on the cylinder set $[C_1]$ is

$$\left\{\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{matrix}\right\} - \left\{\begin{matrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}\right\} = \{B\} - \{B_0\}.$$

The variables on the left hand side are in the basis, and we have just deduced the variable $\{B_0\}$. Therefore we can deduce the variable $\{B\}$, as required.

In a similar way we can deduce all those g-variables containing one 1. Having done that, we can then deduce all the f-variables with one 1 (there are only two of these to deduce, since the four others are basis variables).

For example the variable

$$\left\{\begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}\right\}$$

is deduced by considering the cocycle equation on the cylinder set corresponding to the block

$$\begin{matrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}.$$

So now *all* variables containing at most 1 are considered to be 'known'. These known variables, together with the basis variables containing two 1's, are sufficient to deduce *all* variables containing two 1's.

Continuing in this way we can deduce all variables containing at most four 1's.

The remaining variables (i.e. those containing either five or six 1's) can then be deduced immediately.

For example the f-variable

$$\left\{\begin{matrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{matrix}\right\}$$

is deduced by considering the cocycle equation on the cylinder set corresponding to the block

$$\begin{matrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{matrix},$$

since the other 3 variables all contain at most four 1's.

Section 2·12. Triviality of Hölder Cocycles.

In this section we use our triviality results for locally constant cocycles on semi-safe symbol subshifts X to prove a similar result for Hölder cocycles. The only difference (see Theorem 2·15) is that we assume X satisfies a transitivity condition.

Definition 2·23. *Suppose $X \subseteq A^{\mathbf{Z}^2}$ is a subshift. For any function $\phi : X \to \mathbf{R}$ we define its N^{th}* **variation** *$var_N\phi$ by*

$$var_N\phi = \sup\{|\phi(x) - \phi(y)| : \pi_{S_N}(x) = \pi_{S_N}(y)\}.$$

*For $\alpha \in (0,1]$, we say that ϕ is α-***Hölder** *if there exists some $K > 0$ (the Hölder constant) such that $var_N\phi \leq K\alpha^N$.*

For a continuous cocycle (f,g) on X, we define $var_N(f,g) = \max\{var_N f, var_N g\}$, and say that (f,g) is α-Hölder if there exists $K > 0$ such that $var_N(f,g) \leq K\alpha^N$.

We say that a function (resp. cocycle) is **Hölder** *if it is α-Hölder for some $\alpha \in (0,1]$.*

We remark that the locally constant functions (see Definition 2·15) are precisely those whose N^{th} variation is eventually zero. If a locally constant cocycle (f,g) is of degree N then $var_{N+1}(f,g) = 0$.

It is well known that locally constant functions are uniformly dense in the space of Hölder functions. It follows that locally constant cocycles on *one-dimensional* subshifts are uniformly dense in the space of Hölder cocycles. In two dimensions the analogous result is not so evident. Given a Hölder cocycle (f, g), we can certainly find sequences f_N, g_N of locally constant functions with $f_N \to f$ and $g_N \to g$. However, we must check that this can be done so that each pair (f_N, g_N) is a cocycle. The following result states that if X is a semi-safe symbol subshift then this is possible.

Proposition 2·13. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a semi-safe symbol subshift. The space of locally constant cocycles on X is uniformly dense in the space of Hölder cocycles on X.*

Proof. Without loss of generality we will assume that 0 is a semi-safe symbol of type SouthWest. Let (f, g) be an α-Hölder cocycle on X, with corresponding Hölder constant $K > 0$. We will define a sequence $(f_N, g_N) \in V_N$, and show this sequence converges uniformly to (f, g) as $N \to \infty$. Recall that if $(f_N, g_N) \in V_N$, then f_N, g_N have active coordinates in the rectangles F_N, G_N respectively (see Definition 2·19).

By Theorem 2·12 we know that all cocycles in V_N are trivial. Therefore by defining (f_N, g_N) on its canonical basis variables (see proof of Proposition 2·10), we in fact determine the whole cocycle.

If $\{D\}$ is a basis f_N-variable, then define

$$\{D\} = f_N([D]) = \max_{x \in D} f(x). \tag{2·8}$$

If $\{D'\}$ is the basis g_N-variable (i.e. the all 0's decoration), then define

$$\{D'\} = g_N([D']) = \max_{x \in D'} g(x). \tag{2·9}$$

Now all other variables are given as linear combinations of these basis variables, so we have defined (f_N, g_N). We will show that the *length* of these linear combinations only grows quadratically with N. This is the key to the proof that (f_N, g_N) converges uniformly to (f, g).

First we will show that g_N converges uniformly to g. For any $z \in X$ there exists a unique globally allowed block B of size G_N such that $z \in [B]_{1-N,1-N}$. Let B' be the block of size $2(2N-1) \times 2N$ obtained by placing $2N-1$ columns of 0's to the left of B (possible since 0 is semi-safe of type SouthWest). We will consider the cylinder sets $\sigma^r([B']_{1-N,1-N})$, for $0 \le r \le 2N-1$. We note that the last of these cylinder sets, $\sigma^{2N-1}([B']_{1-N,1-N})$, is contained in $[B]_{1-N,1-N}$, and therefore

$$g_N(\sigma^{2N-1}([B']_{1-N,1-N})) = g_N([B]_{1-N,1-N}). \tag{2·10}$$

For ease of notation we will write $[B]$ and $[B']$ for the cylinder sets $[B]_{1-N,1-N}$ and $[B']_{1-N,1-N}$ respectively. On each of the cylinder sets $\sigma^r([B']_{1-N,1-N})$, we have the cocycle equation (see (2·5)) for (f_N, g_N),

$$g_N(\sigma(\sigma^r[B'])) = g_N(\sigma^r[B']) + f_N(\tau(\sigma^r[B'])) - f_N(\sigma^r[B']). \tag{2·11}$$

We now estimate the difference between $g(z)$ and $g_N(z)$. We successively use the cocycle equation (2·11) to express $g_N(z)$ solely in terms of basis variables.

$$\begin{aligned}
&|g(z) - g_N(z)| = |g(z) - g_N[B]| \\
&= |g(z) - g_N(\sigma(\sigma^{2N-2}[B']))| \quad \text{by (2·10)} \\
&= |g(z) - [g_N(\sigma^{2N-2}[B']) + f_N(\tau(\sigma^{2N-2}[B'])) - f_N(\sigma^{2N-2}[B'])]| \\
&= |g(z) - [\{g_N(\sigma^{2N-3}[B']) + f_N(\tau\sigma^{2N-3}[B']) - f_N(\sigma^{2N-3}[B'])\} \\
&\qquad\qquad + f_N(\tau(\sigma^{2N-2}[B'])) - f_N(\sigma^{2N-2}[B'])]| \\
&\vdots \\
&= |g(z) - [g_N[B'] + \sum_{r=0}^{2N-2} (f_N(\tau\sigma^r[B']) - f_N(\sigma^r[B']))]|.
\end{aligned}$$

We now want to replace the basis variables in the above expression with the values assigned to them in (2·8) and (2·9). First we introduce some notation. Let E'_r be the block obtained from B' by removing r columns from its left. Let E'_r have basepoint $(1-N, 1-N)$ (i.e. this is the bottom left-hand corner of E'_r). Note that $\sigma[E'_r] \subseteq [E'_{r+1}]$. Let C'_r be the $2N \times 2N$ square block obtained from B' by removing r columns from its left and $2N-2-r$ columns from its right. Let C'_r have basepoint $(1-N, 1-N)$. Remove the bottom row of C'_r to obtain the block U'_r with basepoint $(1-N, 2-N)$. Remove the top row of C'_r to obtain the block D'_r with basepoint $(1-N, 1-N)$. Remove the right-hand column of C'_r to obtain the block L'_r with basepoint $(1-N, 1-N)$.

The equality above now becomes

$$|g(z) - g_N(z)| = |g(z) - [\max_{x\in[L'_0]} g(x) + \sum_{r=0}^{2N-2} (\max_{x\in[U'_r]} f(\tau x) - \max_{x\in[D'_r]} f(x))]|. \tag{2·12}$$

Now we can pick some $y \in [B']$. Defining $y_r = \sigma^r y$, we note that each $y_r \in [E'_r]$. Since also $[E'_r]$ is contained in each of $[L'_r]$, $[U'_r]$ and $[D'_r]$, we can successively introduce the points y_r into the expression (2·12). Each such introduction results in an error term $var_N(f,g)$. At each stage we use the fact that (f,g) satisfies the cocycle equation. We have

$$
\begin{aligned}
&|g(z) - g_N(z)| \\
&\le |g(z) - [g(y_0) + f(\tau y_0) - f(y_0) + \sum_{r=1}^{2N-2} (\max_{x\in[U'_r]} f(\tau x) - \max_{x\in[D'_r]} f(x))]| \\
&\qquad + |g(y_0) - \max_{x\in[L'_0]} g(x)| + |f(\tau y_0) - \max_{x\in[U'_0]} f(\tau x)| + |f(y_0) - \max_{x\in[D'_0]} f(x)| \\
&\le |g(z) - [g(\sigma y_0) + \sum_{r=1}^{2N-2} (\max_{x\in[U'_r]} f(\tau x) - \max_{x\in[D'_r]} f(x))]| + 3var_N(f,g) \\
&\le |g(z) - [g(y_1) + f(\tau y_1) - f(y_1) + \sum_{r=2}^{2N-2} (\max_{x\in[U'_r]} f(\tau x) - \max_{x\in[D'_r]} f(x))]| \\
&\qquad + |f(\tau y_0) - \max_{x\in[U'_0]} f(\tau x)| + |f(y_0) - \max_{x\in[D'_0]} f(x)| + 3var_N(f,g) \\
&\le |g(z) - [g(\sigma y_1) + \sum_{r=2}^{2N-2} (\max_{x\in[U'_r]} f(\tau x) - \max_{x\in[D'_r]} f(x))]| + 2var_N(f,g) + 3var_N(f,g) \\
&\ \vdots \\
&\le |g(z) - g(\sigma y_{2N-2})| + 2(2N-2)var_N(f,g) + 3var_N(f,g) \\
&\le var_N(f,g) + (4N-1)var_N(f,g) \qquad \text{since } \sigma(y_{2N-2}) \in [E'_{2N-1}] = [B] \\
&= 4N var_N(f,g) \\
&\le 4NK\alpha^N \to 0 \quad \text{as } N \to \infty.
\end{aligned}
$$

Therefore we have shown that $\|g - g_N\|_\infty \to 0$ as $N \to \infty$, and that the convergence is exponentially fast.

The proof that $\|f - f_N\|_\infty \to 0$ as $N \to \infty$ is similar. This time we pick an arbitrary $z \in X$, and define B to be the block of size F_N such that $z \in [B]$. If $f_N[B]$ is a basis variable then we clearly have that

$$|f(z) - f_N(z)| = |f(z) - f_N[B]| = |f(z) - \max_{x \in [B]} f(x)| \leq var_N(f,g) \leq K\alpha^N \to 0.$$

If $f_N[B]$ is not a basis variable then we build the block B' by adding $2N-1$ rows of 0's to the bottom of B. As before we go along B' and use the cocycle equation on each $2N \times 2N$ square block. Eventually we obtain an expression for $|f(z) - f_N(z)|$ which is solely in terms of known variables. The difference is that all but one of these known variables is a g_N-variable (rather than a basis variable). We again pick some $y \in [B']$, and successively introduce the iterates $\tau^r y$ into the expression for $|f(z) - f_N(z)|$. Each such introduction results in an error term $4N var_N(f,g)$ (compare to the smaller error terms involved in the estimate of $|g(z) - g_N(z)|$). Eventually we obtain that

$$|f(z) - f_N(z)| \leq [8N(2N-1)+2] var_N(f,g) \leq [8N(2N-1)+2]K\alpha^N \to 0 \text{ as } N \to \infty.$$

So we have shown that $\|f - f_N\|_\infty \to 0$ as $N \to \infty$, and that the convergence is exponentially fast. Combining this with the analogous result for g_N, we see that $(f_N, g_N) \to (f,g)$ uniformly, where the convergence is exponentially fast. □

Remark. Proposition 2·13 in fact shows that locally constant cocycles are dense in the set of cocycles (f,g) whose variation $var_N(f,g)$ decays at rate $o(1/N^2)$ as $N \to \infty$.

We would like to combine Proposition 2·13 with Theorem 2·12 to show that all real-valued Hölder cocycles are trivial. We will impose a transitivity condition on X, which allows us to make use of the following theorem due to A. Livsic.

Theorem 2·14. *(Livsic, [32]) Suppose (Y,σ) is a topologically transitive (one dimensional) subshift. Suppose $f : X \to \mathbf{R}$ is Hölder, and satisfies $\sum_{r=0}^{m-1} f(\sigma^r y) = 0$ for all $y \in Y$ with $\sigma^m y = y$. Then there exists a Hölder function $u : X \to \mathbf{R}$ such that $f = u \cdot \sigma - u$.*

□

Theorem 2·15. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a semi-safe symbol subshift. Suppose some point in X has a dense σ-orbit and a dense τ-orbit. Then every Hölder cocycle (f, g) on X is a trivial cocycle (with Hölder transfer function).*

Proof. Let (f, g) be a Hölder cocycle on X. By Proposition 2·13 we can choose a sequence (f_N, g_N) of locally constant cocycles converging uniformly to (f, g). Let $\underline{x} \in X$ be the fixed point decorated solely by the semi-safe symbol, and let $c_f = f(\underline{x})$, $c_g = g(\underline{x})$. Our construction in Proposition 2·13 ensures that $f_N(\underline{x}) = c_f$ and $g_N(\underline{x}) = c_g$, for all $N \geq 1$. By Theorem 2·12 we know that each (f_N, g_N) is trivial. So there exist locally constant transfer functions $h_N : X \to \mathbf{R}$ such that

$$f_N = h_N \cdot \sigma - h_N + c_f \tag{2·13}$$

$$g_N = h_N \cdot \tau - h_N + c_g \ . \tag{2·14}$$

Suppose $z \in X$ has a dense σ-orbit and a dense τ-orbit. Since transfer functions are unique up to an additive constant (see Corollary 2·6), we may assume that $h_N(z) = b$, say, for all $N \geq 1$. Now equation (2·13) gives us $\sum_{r=0}^{m-1} [f_N(\sigma^r y) - c_f] = 0$ for all $y \in X$ satisfying $\sigma^m y = y$. Letting $N \to \infty$ gives that $\sum_{r=0}^{m-1} [f(\sigma^r y) - c_f] = 0$ for all $y \in X$ satisfying $\sigma^m y = y$. Since $f - c_f$ is Hölder, Theorem 2·14 guarantees us a Hölder function u satisfying

$$f - c_f = u \cdot \sigma - u \ . \tag{2·15}$$

By a similar argument we know there exists a Hölder function v satisfying

$$g - c_g = v \cdot \tau - v \ . \tag{2·16}$$

Since transfer functions are unique up to an additive constant, we may assume that $u(z) = v(z) = b$. To prove that (f, g) is trivial, we must show that $u = v$. In fact we claim they are both equal to $\lim_{N\to\infty} h_N$.

From equation (2·13) we have

$$h_N(\sigma^m(z)) = b + \sum_{r=0}^{m-1} [f_N(\sigma^r z) - c_f] \quad \text{for all } m \geq 0.$$

Letting $N \to \infty$ gives, for all $m \geq 0$,

$$\begin{aligned}\lim_{N\to\infty} h_N(\sigma^m(z)) &= b + \sum_{r=0}^{m-1} [f(\sigma^r z) - c_f] \\ &= b + u(\sigma^m z) - u(z) \quad \text{by } (2{\cdot}15) \\ &= u(\sigma^m z).\end{aligned}$$

The same equality holds for $m < 0$. Thus u and $\lim_{N\to\infty} h_N$ agree on the σ-orbit of z. But this orbit is dense, and u is continuous, therefore $\lim_{N\to\infty} h_N$ exists everywhere, and is equal to u. Applying a similar argument to the τ-orbit of z, we see that $\lim_{N\to\infty} h_N = v$.

Therefore equations (2·15) and (2·16) together show that (f, g) is trivial, with Hölder transfer function $\lim_{N\to\infty} h_N$. $\square$

The hypotheses of Theorem 2·15 are satisfied by any safe symbol subshift of finite type, as the following lemma demonstrates.

Lemma 2·16. *Suppose $X \subseteq A^{\mathbb{Z}^2}$ is a safe symbol subshift of finite type. Then there exists a point $z \in X$ whose σ-orbit and τ-orbit are both dense in X.*

Proof. Let (B_i) be an enumeration of all globally allowed square blocks of odd side length, such that the size of B_i is at least as big as the size of B_{i-1}. We must construct a point $z \in X$ which contains all of these blocks B_i.

We start the construction of z by laying the B_i, in order, along the positive horizontal axis of $\mathbb{Z}^2$, leaving a 'gap' (i.e. at least one point of $\mathbb{Z}^2$) between consecutive blocks. We do this in a symmetric way, so that half of each block lies above the axis and half lies below it. Similarly we lay the B_i, in order, along the positive vertical axis, again leaving a 'gap' between consecutive blocks, and such that half of each block lies to the left of the axis and half lies to the right.

Next we simply decorate all remaining coordinates of $\mathbb{Z}^2$ with the safe symbol to obtain a point $z \in X$. Since every globally allowed block appears in the positive horizontal (resp. vertical) direction of z, the forward σ-orbit (resp. τ-orbit) of z will visit each non-empty cylinder set of X. Thus both these orbits are dense in X. $\square$

Corollary 2·17. *Let $X \subseteq A^{\mathbf{Z}^2}$ be a safe symbol subshift of finite type. Then every Hölder cocycle (f, g) on X is a trivial cocycle (with Hölder transfer function).*

Proof. This is immediate from Theorem 2·15 and Lemma 2·16. □

Section 2·13. Locally (Residually Finite) Group-valued Cocycles.

Up until now we have restricted our attention to cocycles on a subshift X taking values in the additive group $\mathbf{R}$. We used the field structure of $\mathbf{R}$ to define the finite dimensional vector spaces V_N and V'_N (see Definition 2·21). We then used a dimension argument to show that, if X is a semi-safe symbol subshift, then V_N and V'_N are equal. This allowed us to deduce (Theorem 2·12) that all locally constant cocycles are trivial.

In this section we investigate cocycles taking values in more general groups G, which we always write multiplicatively. We will prove that if G is locally (residually finite), then all locally constant cocycles are trivial. We remark that the class of locally (residually finite) groups is large. For example it contains all abelian groups, metabelian groups, locally (polycyclic-by-finite) groups, linear groups, matrix groups over finitely generated integral domains, and all free groups. In particular, certain groups without a doubly invariant metric (for example, general linear groups) are locally (residually finite). Such groups were not covered by the techniques in Schmidt [63].

Throughout this section we shall assume that $X \subseteq A^{\mathbf{Z}^2}$ is a semi-safe symbol subshift.

First we amend some of our definitions of cohomology, to allow for the case where G is non-abelian. The cocycle $F : \mathbf{Z}^2 \times X \to G$ (cf. remarks after Definition 2·18) must satisfy (see [63]) the equation

$$F(m + m', n + n', x) = F(m', n', \sigma^m \tau^n(x))F(m, n, x).$$

Writing the cocycle as the pair of generating functions (f, g), where $f(x) = F(1, 0, x)$, $g(x) = F(0, 1, x)$, we have the following definition.

Definition 2·24. *Let G be a group, and $X \subseteq A^{\mathbf{Z}^2}$ a subshift. Suppose $f, g : X \to G$ are (locally constant) functions. The pair of functions (f, g) is said to be a (locally constant)*

G-valued cocycle *on X if for all $x \in X$,*

$$f(\tau x)g(x) = g(\sigma x)f(x). \tag{2·17}$$

Definition 2·25. *A locally constant G-valued cocycle (f, g) on a subshift X is said to be* **trivial** *if there exist constants $c_f, c_g \in G$ and a function $h : X \to G$ such that for all $x \in X$,*

$$f(x) = h(\sigma x)c_f h(x)^{-1} \quad \text{and} \quad g(x) = h(\tau x)c_g h(x)^{-1}. \tag{2·18}$$

Such an h is called a **transfer function**.

Remark. If X is a semi-safe symbol subshift, then it contains a fixed point $\underline{x}$ (see Lemma 2·2). If (f, g) is a G-valued cocycle on X then (2·17) implies that $f(\underline{x})$ and $g(\underline{x})$ commute. Moreover, if (f, g) is trivial, then (2·18) implies that the constants c_f, c_g also commute.

The following result is a generalisation of Corollary 2·6.

Lemma 2·18. *Let X be a semi-safe symbol subshift, and suppose that (f, g) is a trivial locally constant G-valued cocycle on X, for some group G. Suppose $h, h' : X \to G$ are both transfer functions for (f, g). Then there exists $b \in G$ such that $h(z) = h'(z)b$ for all $z \in X$.*

Proof. We may assume that the semi-safe symbol is of type SouthWest. Since h, h' must themselves be locally constant, we may assume that their active coordinates both lie in the square S_N (see Definition 2·4) for some $N \geq 1$. Suppose $\underline{x} \in X$ is the fixed point decorated solely by the semi-safe symbol. Then there is a dense subset $\Lambda \subset X$ of points z such that $(\sigma^{-1}\tau^{-1})^i(z) \to \underline{x}$ as $i \to \infty$. So for each $z \in \Lambda$ there exists $M_z \in \mathbb{N}$ such that if $i \geq M_z$ then the square block $\pi_{S_N}((\sigma^{-1}\tau^{-1})^i(z))$ is decorated solely by the semi-safe symbol. It follows that

$$h((\sigma^{-1}\tau^{-1})^i(z)) = h(\underline{x}) \quad \text{and} \quad h'((\sigma^{-1}\tau^{-1})^i(z)) = h'(\underline{x}) \quad \text{for all } i \geq M_z. \tag{2·19}$$

Now since (f,g) is trivial, there are commuting pairs of constants c_f, c_g and d_f, d_g such that

$$f(z) = h(\sigma z)c_f h(z)^{-1} \quad , \quad g(z) = h(\tau z)c_g h(z)^{-1},$$

$$f(z) = h'(\sigma z)d_f h'(z)^{-1} \quad , \quad g(z) = h'(\tau z)d_g h'(z)^{-1}.$$

This implies that

$$h(\sigma^m \tau^n z)c_f^m c_g^n h(z)^{-1} = h'(\sigma^m \tau^n z)d_f^m d_g^n h'(z)^{-1} \quad \text{for all } (m,n) \in \mathbf{Z}^2. \tag{2·20}$$

Setting $m = n = -M_z$, and writing $\gamma = c_f c_g$, $\delta = d_f d_g$, we obtain

$$h(\underline{x})\gamma^{-M_z} h(z)^{-1} = h'(\underline{x})\delta^{-M_z} h'(z)^{-1}.$$

Rearranging this equation gives

$$h'(z)^{-1} h(z) = \delta^{M_z} h'(\underline{x})^{-1} h(\underline{x})\gamma^{-M_z}. \tag{2·21}$$

In equation (2·20) we can also set $m = n = -(M_z + 1)$, and by the same process we obtain

$$h'(z)^{-1} h(z) = \delta^{M_z+1} h'(\underline{x})^{-1} h(\underline{x})\gamma^{-(M_z+1)}. \tag{2·22}$$

Equating (2·21) and (2·22), then left-multiplying by δ^{-M_z} and right-multiplying by γ^{M_z}, gives us

$$\delta h'(\underline{x})^{-1} h(\underline{x})\gamma^{-1} = h'(\underline{x})^{-1} h(\underline{x}),$$

and by induction we deduce that

$$\delta^j h'(\underline{x})^{-1} h(\underline{x})\gamma^{-j} = h'(\underline{x})^{-1} h(\underline{x}) \quad \text{for all } j \geq 0.$$

In particular, for all $z \in \Lambda$ we have

$$\delta^{M_z} h'(\underline{x})^{-1} h(\underline{x})\gamma^{-M_z} = h'(\underline{x})^{-1} h(\underline{x}). \tag{2·23}$$

Substituting (2·23) into (2·21) gives

$$h'(z)^{-1} h(z) = h'(\underline{x})^{-1} h(\underline{x}) \quad \text{for all } z \in \Lambda.$$

So the continuous function $z \mapsto h'(z)^{-1}h(z)$ is constant on the dense set Λ, and is therefore constant on all of X. If we call the constant value b, then the result follows. $\square$

As in Definition 2·21, we can define the set $V_N(X)$ of G-valued cocycles of degree N, and the set $V_N'(X)$ of trivial G-valued cocycles of degree N. If G is (the additive group of) a field, then these sets are both vector spaces over G. In fact, it is useful to think of them as subspaces of the vector space $G^{|F_N|+|G_N|} = G^{4N(2N-1)}$. If G is abelian, then V_N and V_N' are both subgroups of the group $G^{4N(2N-1)}$. If G is non-abelian, then V_N and V_N' are both subsets of $G^{4N(2N-1)}$, but do not themselves carry a group structure.

As in Lemma 2·7, if $(f,g) \in V_N$ then each of the four functions in the cocycle equation (2·17) has active coordinates in the square T_N. Therefore the $|\pi_{F_N}(X)|$ f-variables and $|\pi_{G_N}(X)|$ g-variables satisfy a system of $|\pi_{T_N}(X)|$ equations. Although these equations are no longer linear (such a notion is meaningless for general groups G), this is unimportant. What is important is that in each equation (2·17) there are just two f-variables and two g-variables, with one of each type on either side. Multiplication by an appropriate inverse means that each of these variables can be expressed as a *word* in the other three variables (if G is abelian then this is just a $\mathbb{Z}$-linear combination, with coefficients ± 1).

We would like to use the arguments in the proof of Proposition 2·10, where we used the system of cocycle equations to express all variables in terms of a small set of basis variables, to show that the sets $V_N(X)$ and $V_N'(X)$ are equal.

In fact it will be more convenient to partition each V_N, V_N' into sets of cocycles which agree on the fixed point decorated solely by the semi-safe symbol. We make the following definition.

Definition 2·26. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a semi-safe symbol subshift, and suppose $\underline{x} \in X$ is the fixed point decorated solely by the semi-safe symbol. Let G be a group, and suppose $i, j \in G$ satisfy $ij = ji$. Define*

$$V_N(i,j) = \{(f,g) \in V_N : f(\underline{x}) = i,\ g(\underline{x}) = j\},$$

$$V_N'(i,j) = \{(f,g) \in V_N' : f(\underline{x}) = i,\ g(\underline{x}) = j\}.$$

Again it is useful to think of $V_N(i,j)$, $V'_N(i,j)$ as subsets of $G^{4N(2N-1)}$. As was the case for V_N, V'_N, the algebraic structure of $V_N(i,j)$, $V'_N(i,j)$ depends on that of G. If G is (the additive group of) a field then $V_N(i,j)$, $V'_N(i,j)$ are vector subspaces of $G^{4N(2N-1)}$. If G is an abelian group then $V_N(i,j)$, $V'_N(i,j)$ are subgroups of $G^{4N(2N-1)}$, while if G is non-abelian, then $V_N(i,j)$, $V'_N(i,j)$ do not carry a group structure. We would like some notion of the 'dimension' of the sets $V_N(i,j)$ and $V'_N(i,j)$. For the moment we talk in rather loose terms about the 'degrees of freedom' we have in choosing cocycles from these sets. A degree of freedom corresponds to an unrestricted choice of an element of G. The number of degrees of freedom can be interpreted as the number of *freely varying parameters*. Later we make this notion more precise.

Lemma 2·19. *Let G be a group, and $X \subseteq A^{\mathbb{Z}^2}$ a semi-safe symbol subshift. For any $i,j \in G$ satisfying $ij = ji$, we have $|\pi_{S_N}(X)| - 1$ degrees of freedom in choosing a cocycle from $V'_N(i,j)$.*

Proof. The transfer function h of a trivial cocycle is only unique up to a constant (see Lemma 2·18). Once we have specified $h(\underline{x})$, where $\underline{x} \in X$ is the fixed point decorated solely by the semi-safe symbol, equation (2·18) gives us the constants $c_f, c_g \in G$ in terms of i,j, and $h(\underline{x})$. Since i,j commute, then so do c_f, c_g. We have complete freedom in the choice of the remaining $|\pi_{S_N}(X)| - 1$ h-variables, and now we have completely specified our trivial cocycle. $\square$

Proposition 2·20. *Let G be a group, and $X \subseteq A^{\mathbb{Z}^2}$ a semi-safe symbol subshift. For any $i,j \in G$ satisfying $ij = ji$, we have $|\pi_{S_N}(X)| - 1$ degrees of freedom in choosing a cocycle from $V_N(i,j)$.*

Proof. The argument is exactly the same as was used in Propositions 2·10 and 2·11. We choose the same basis of $|\pi_{S_N}(X)| + 1$ variables, and use the system of cocycle equations to express *every* variable as a word in the basis variables. The key fact is that each cocycle equation can be rearranged so that one variable is expressed in terms of the other three.

Let $\underline{x} \in X$ denote the fixed point decorated solely by the semi-safe symbol. Since the values of $f(\underline{x})$ and $g(\underline{x})$ are fixed to be i and j respectively, the same is true of the corresponding two variables in the basis. This leaves $|\pi_{S_N}(X)|-1$ basis variables. Since all remaining variables are f-variables, there are no commutativity relations between them, so they are each free to range over all of G. $\square$

The following two facts hold for any $i,j \in G$ with $ij = ji$. The first fact follows straight from Definition 2·26, while the second fact is a consequence of Lemma 2·19 and Proposition 2·20.

1. $V'_N(i,j) \subseteq V_N(i,j)$.
2. $V_N(i,j)$ and $V'_N(i,j)$ are subsets of $G^{4N(2N-1)}$ which are parametrised by the same number of freely-varying parameters.

Definition 2·27. *A group G is called* **good** *if for all $N \geq 1$, and all $i,j \in G$ with $ij = ji$, the conditions 1 and 2 above together imply that $V_N(i,j) = V'_N(i,j)$.*

Proposition 2·21. *Let $X \subseteq A^{\mathbb{Z}^2}$ be a semi-safe symbol subshift. Then G is a good group if and only if all G-valued locally constant cocycles on X are trivial.*

Proof. Since G is good, then Lemma 2·19 and Proposition 2·20 imply that $V_N(i,j) = V'_N(i,j)$ for all $i,j \in G$ with $ij = ji$. Taking the union over all such i,j implies that $V_N = V'_N$. Taking the union over all $N \geq 1$, we see that all locally constant G-valued cocycles on X are trivial.

Conversely, if G is not good then we can find some $N \geq 1$, and some $i,j \in G$ with $ij = ji$, such that $V'_N(i,j)$ is *strictly* contained in $V_N(i,j)$. Therefore there exists a non-trivial cocycle. $\square$

If G is the additive group of a field, then G is certainly good, by linear algebra, since $V_N(i,j)$, $V'_N(i,j)$ are vector subspaces of the same dimension. Any *finite* group G is also good, since in this case $V_N(i,j)$, $V'_N(i,j)$ have the same number (namely $|G|^{|\pi_{S_N}(X)|-1}$) of elements. To find a larger class of good groups, we will generalise our problem so as to omit any reference to the specific sets $V_N(i,j)$, $V'_N(i,j)$. First we introduce some notation, so as to put the notion of 'degrees of freedom' in a more rigorous context. I am grateful to

Giovanni Cutolo for explaining to me the abstract group theory in the remainder of this section.

Fix a free group F of rank m. Let $x_1, \ldots, x_m$ be a basis for F. Let $w = (w_1, \ldots, w_n)$ be an n-tuple of elements (words) of F, such that $w_1, \ldots, w_n$ generate F (so $n \geq m$).

Define $w^* : G^m \to G^n$ by

$$w(g_1, \ldots, g_m) = (w_1(g_1, \ldots, g_m), \ldots, w_n(g_1, \ldots, g_m)).$$

Define the set $S(w, G) =$ Image $(w^*) \subset G^n$. (The correspondence with our original notation is given by $m = |\pi_{S_N}(X)| - 1$, $n = 4N(2N-1)$. The set $S(w,G)$ will represent either $V_N(i,j)$ or $V'_N(i,j)$).

Let us fix some further notation. Since the elements w_i generate F, there are words $\hat{w}_1, \ldots, \hat{w}_m$ on n symbols such that $\hat{w}_i(w) = x_i$ for each $i \in \{1, \ldots, m\}$.

Define $\hat{w} : G^n \to G^m$ by

$$\hat{w}(g_1, \ldots, g_n) = (\hat{w}_1(g_1, \ldots, g_n), \ldots, \hat{w}_m(g_1, \ldots, g_n)).$$

Note that $\hat{w} \circ w^* : G^m \to G^m$ is the identity map. In particular, w^* is injective.

If we choose another n-tuple $v = (v_1, \ldots, v_n)$ of generators of F, then the set $S(v, G)$ and the maps v^*, $\hat{v}$ can be defined similarly.

Notice that $S(w, G)$ is contained in $S(v, G)$ if and only if $w^* = v^* \circ \hat{v} \circ w^*$. This is because if $g \in G^m$ satisfies $w^*(g) = v^*(h)$ for some $h \in G^m$, then $h = \hat{v} \circ v^*(h) = \hat{v} \circ w^*(g)$. Applying v^* to both sides gives $v^*(h) = v^* \circ \hat{v} \circ w^*(g)$, which implies $w^*(g) = v^* \circ \hat{v} \circ w^*(g)$.

So for fixed n-tuples w and v, the class $Y(w, v)$ of groups G for which $S(w, G)$ is contained in $S(v, G)$ is defined by a set of equations between values of words. Thus $Y(w, v)$ is a variety of groups (see Robinson [57], page 56). Let $Z(w, v)$ be the class of groups G such that $S(w, G)$ is contained in $S(v, G)$ if and only if the reverse inclusion holds. Let Z denote the intersection of $Z(w, v)$ over all possible n-tuples w, v.

Definition 2·28. *We call a group* **very good** *if it belongs to the class Z.*

Remark. If G is a very good group, then it is certainly good (see Definition 2·27), since the sets $S(w, G)$, $S(v, G)$ are generalisations of $V_N(i,j)$, $V'_N(i,j)$. We are not sure whether or not every good group is very good. The class Z' of good groups would be the intersection of certain classes $Z(w, v)$, where we take the intersection over all possible n-tuples w, v of words which arise when solving the system of cocycle equations. There may be a restriction on the type of words which can arise in this way, in which case the class Z' could be strictly larger than Z. Nevertheless, the class Z is already rather large, as we now demonstrate.

Definition 2·29. *Suppose $\mathcal{X}$ is some property of groups. We say a group G is* **residually** *$\mathcal{X}$ if for all $1 \neq g \in G$ there exists a normal subgroup N_g such that $g \notin N_g$ and G/N_g satisfies $\mathcal{X}$.*

Definition 2·30. *Suppose $\mathcal{X}$ is some property of groups. We say a group G is* **locally (residually $\mathcal{X}$)** *if all of its finitely generated subgroups are residually $\mathcal{X}$.*

Proposition 2·22. *Every locally (residually finite) group is a very good group.*

Proof. Fix n-tuples w, v as above. If G is a finite group then, with the same notation as before, we have $|S(w, G)| = |S(v, G)| = |G|^m$. Thus $S(w, G) \subset S(v, G)$ if and only if $S(v, G) \subset S(w, G)$. So any finite group belongs to the class $Z(w, v)$.

Now since $Y(w, v)$ and $Y(v, w)$ are both varieties, it follows that any group which is locally (residually in $Z(w, v)$) is itself in $Z(w, v)$ (see Robinson [57], page 57).

So any group which is locally (residually finite) belongs to $Z(w, v)$. Since w, v were arbitrary, the result follows. □

Theorem 2·23. *Let $X \subset A^{\mathbf{Z}^2}$ be a semi-safe symbol subshift, and suppose G is a locally (residually finite) group. Then all locally constant G-valued cocycles on X are trivial.*

Proof. Since G is locally (residually finite), then by Proposition 2·22 it is a very good group. In particular G is a good group, so by Proposition 2·21 we deduce the required result. □

The class of locally (residually finite) groups is discussed in Chapter 9 of Robinson [56]. The following classes of groups are all locally (residually finite).

(a) All abelian groups.

(b) All metabelian groups (i.e. those soluble groups of derived length at most two). This class includes all abelian groups.

(c) All locally (polycyclic-by-finite) groups. This class contains all metabelian groups. A group G is *polycyclic* if there is a chain of normal subgroups $1 = G_0 \lhd G_1 \lhd \cdots \lhd G_n = G$ such that each quotient G_{i+1}/G_i is cyclic. A group is *polycyclic-by-finite* if it has a polycyclic normal subgroup of finite index. A group is *locally (polycyclic-by-finite)* if all its finitely generated subgroups are polycyclic-by-finite. The fact that this class of groups is locally (residually finite) is a consequence of the Jategaonkar-Roseblade Theorem (see Theorem 6.6 in Passman [51]).

(d) All groups of matrices over finitely generated integral domains are residually finite, and hence locally (residually finite) (see chapter 4 of Wehrfritz [74]).

(e) All free groups (see Robinson [57], page 158).

The following results are of particular interest, as the triviality of cocycles taking values in general linear groups does not follow from the results in Schmidt [63]. This is because $GL(n, \mathbf{C})$ does not admit a doubly invariant metric (see Hewitt & Ross [23], page 78).

Proposition 2·24. *If F is a field, then any matrix group $M(F)$ over F is locally (residually finite).*

Proof. Suppose H is a finitely generated subgroup of $M(F)$. Choose a finite set B of generators of H, and let C be the set of all elements of F that appear as entries of elements of B. Let R be the subring of F generated by C. Then H can be regarded as a group of matrices over the finitely generated integral domain R, and is therefore residually finite by (d) above. Thus $M(F)$ is locally (residually finite). $\square$

Corollary 2·25. *Suppose $X \subseteq A^{\mathbf{Z}^2}$ is a semi-safe symbol subshift. For every $n \geq 1$, all locally constant $GL(n, \mathbf{C})$-valued cocycles on X are trivial.*

Proof. Setting $F = \mathbf{C}$ in Proposition 2·24 shows that $GL(n, \mathbf{C})$ is locally (residually finite). The result follows from Theorem 2·23. □

Remark. We do not know of any group G for which there exist non-trivial locally constant G-valued cocycles on a semi-safe symbol subshift X. It is possible that the property of local (residual finiteness) does not completely characterise the class of good groups (see Definition 2·27). In fact we do not know whether this property completely characterises the class of very good groups (see Definition 2·28), nor whether the class of very good groups is strictly contained in the class of good groups.

Chapter 3. Barycentres Of Invariant Measures For The Doubling Map

Section 3·1. Introduction.

In this chapter we consider the set $\mathcal{M}$ of Borel probability measures invariant under the doubling map of the circle. To each measure we assign a *barycentre*, which is just the integral of the identity function around S^1. We define Ω to be the set of all such barycentres. We study the geometry of Ω, and examine its relationship with $\mathcal{M}$. Elementary considerations show that Ω is a compact, convex subset of the unit disc in $\mathbf{C}$, and is symmetric about the real axis. We present some numerical work regarding the nature of the boundary $\partial\Omega$, and on the basis of this we formulate several conjectures, as yet unproved. Our main conjecture (see Conjecture III in §3·11) is that each point on $\partial\Omega$ is the barycentre of a unique invariant measure, and that such measures are always concentrated on the closure of an *ordered orbit* (i.e. an orbit completely contained in some semi-circle). In particular, such measures have zero entropy. We conjecture further that $\partial\Omega$ is non-differentiable at a countable dense set of points, and that these points correspond to measures supported on *periodic* ordered orbits. The conjectured non-differentiability of $\partial\Omega$ represents the worst possible regularity of the boundary of a planar convex figure (see §3·8).

Conjecture II in §3·10 is equivalent to Conjecture III, and is stated in terms of a certain parametrisation of $\partial\Omega$. We conjecture that this parametrisation is locally constant on a set of full Lebesgue measure, and that each interval of local constancy corresponds to the barycentre of a measure supported on a *periodic* ordered orbit. The parametrisation is such that these intervals of local constancy would correspond to points of non-differentiability of $\partial\Omega$.

We reformulate these conjectures in terms of a certain one-parameter family f_θ of analytic functions (see Conjecture I in §3·10). This conjecture states that the (uniformly) strictly maximal periodic orbits of the family f_θ are precisely the ordered orbits, and that every *periodic* ordered orbit is uniformly strictly maximal for an open interval of parameter values. We believe that Conjecture I is equivalent to Conjectures II and III, but are unable to prove the full extent of this equivalence.

We compare the status of Conjectures I, II and III with an open problem of Pollicott and Sharp (see Conjecture IV in §3·12). We show (Propositions 3·38, 3·40 and 3·41) that Conjectures I, II and III are incompatible with Conjecture IV.

To avoid confusion between proved results and conjectured results, we confine the various conjectures and their corollaries to §3·10, §3·11, and §3·12.

In §3·2 we start with the observation (Lemma 3·3) that every extremal point of Ω has a realisation as the barycentre of an ergodic measure. We note that barycentres of finitely supported measures are dense in Ω, and in §3·4 we compute certain of these barycentres. This gives an idea of the shape of Ω, and in Appendix D we plot an approximation to the boundary $\partial\Omega$. The points on this plot, together with the symbolic codes of the corresponding supports, are listed in Appendix C.

A number-theoretic interlude in §3·5 demonstrates that the origin in $\mathbb{C}$ is the barycentre of infinitely many invariant measures.

In §3·9 we prove (see Proposition 3·26) that every interior point w of Ω is the barycentre of a particular kind of equilibrium state, and that this equilibrium state maximises entropy over all measures whose barycentre is w. This entropy is positive, and we derive a formula for it in terms of pressure (see Corollary 3·27). Positive entropy in the interior of Ω is in marked contrast to the conjectured zero entropy on $\partial\Omega$.

The results about equilibrium states rely on the characterisation of $\partial\Omega$ in terms of a positive functional Q and a certain family of trigonometric functions. The functional Q is studied in a more abstract context in §3·3, where we also introduce the related concepts of maximal measures and (uniformly) strictly maximal orbits. Using a result of Atkinson [4] we prove that a Hölder function with a uniformly strictly maximal periodic orbit has a unique maximal measure. In §3·10 we indicate how this result relates to the conjectured points of non-differentiability of $\partial\Omega$. A related result (Proposition 3·14) of Z. Coelho [12] is used to prove (Proposition 3·23) that the barycentre of a fully supported measure cannot lie on $\partial\Omega$. We remark that if Conjecture III is true then in fact any barycentre on $\partial\Omega$ will correspond to a measure whose support has zero Hausdorff dimension.

In §3·6 we introduce various equivalent definitions of an ordered orbit. We prove that the restriction of the doubling map to the closure of such an orbit is uniquely ergodic. In §3·7 we state a result due to Bullett & Sentenac [10] which describes the mapping of parametrised semi-circles to their associated ordered orbit. This mapping is locally constant on a set of full Lebesgue measure, and the intervals of local constancy correspond to the periodic ordered orbits. This result lends credence to our conjectures in §3·10 and §3·11.

Section 3·2. Definitions and Preliminary Results.

We will use two different models of the circle. K will denote the additive circle $[0, 1)$, with addition defined modulo one, and the usual distance function $d(\cdot, \cdot)$.

$S^1 = \{z \in \mathbb{C} : |z| = 1\}$ will denote the multiplicative circle.

Let the **doubling map** $T : K \to K$ be defined by $T(x) = 2x \pmod 1$.

Let $T' : S^1 \to S^1$ be defined by $T'(z) = z^2$.

Let $\mathcal{M}$ (resp. $\mathcal{M}'$) denote the set of T-invariant (resp. T'-invariant) Borel probability measures on K (resp. S^1).

The homeomorphism $\psi(x) = e^{2\pi i x}$ gives a topological conjugacy between T and T', and induces a one-to-one correspondence between $\mathcal{M}$ and $\mathcal{M}'$. For ease of notation we will write T for both maps, and $\mathcal{M}$ for both sets of invariant measures. If $x \in K$ (resp. $z \in S^1$) then we let $\mathcal{O}(x)$ (resp. $\mathcal{O}(z)$) denote its forward orbit under T.

We recall (see Walters [72]) that $\mathcal{M}$ is a convex set, and is compact in the weak* topology. The extremal points of $\mathcal{M}$ are precisely the ergodic measures. T has a countable infinity of periodic points, and these are dense in the circle. Let

$$\mathrm{Fix}(n) = \left\{ \frac{k}{(2^n - 1)} : 0 \le k \le 2^n - 2 \right\}$$

denote those points of (not necessarily least) period n under T. If $x \in \mathrm{Fix}(n)$ then we can concentrate a unique T-invariant Borel probability measure on $\mathcal{O}(x)$ by setting

$$\mu_x = \frac{1}{n}\sum_{r=0}^{n-1} \delta_{T^r x} ,$$

where δ_y denotes Dirac measure concentrated on the point y. The corresponding measure on S^1, also denoted μ_x, is concentrated on $\mathcal{O}(e^{2\pi i x})$. Define

$$\mathcal{M}_{pp} = \{\mu_x : x \in \mathrm{Fix}(n), n \geq 1\}$$

to be the set of **periodic point measures**, and note that all such measures are ergodic with zero entropy. Moreover, $\mathcal{M}_{pp}$ is weak* dense in $\mathcal{M}$ (see Denker, Grillenberger & Sigmund [14], page 196).

Definition 3·1. *For $\mu \in \mathcal{M}$ we define its* **barycentre** *to be the complex integral $\int_{S^1} z \, d\mu(z)$.*

Our convention is that barycentres will be denoted $\int_{S^1}$, while integrals of real-valued functions (see §3·3) will just be denoted $\int$.

It will sometimes be useful to identify the complex plane $\mathbf{C}$ with $\mathbf{R}^2$. In particular, if $z_1, z_2 \in \mathbf{C}$ with $z_j = x_j + iy_j$, then we let $< z_1, z_2 >= x_1x_2 + y_1y_2$ denote the Euclidean inner product. We will study the following subset of $\mathbf{C}$.

Definition 3·2. *We define*

$$\Omega = \left\{ \int_{S^1} z \, d\mu(z) : \mu \in \mathcal{M} \right\}$$

to be the **barycentre set**.

Note that Lebesgue measure l is invariant under T, so $0 = \int_{S^1} z \, dl(z) \in \Omega$.

Note that if $x \in \mathrm{Fix}(n)$ then the barycentre of μ_x is a finite trigonometric sum:

$$\begin{aligned}\int_{S^1} z \, d\mu_x(z) &= \frac{1}{n}\sum_{r=0}^{n-1} e^{2\pi i 2^r x} \\ &= \frac{1}{n}\left(\sum_{r=0}^{n-1} \cos(2\pi 2^r x) + i \sum_{r=0}^{n-1} \sin(2\pi 2^r x)\right).\end{aligned}$$

We let

$$\Omega_{pp} = \left\{ \frac{1}{n}\sum_{r=0}^{n-1} e^{2\pi i 2^r x} : x \in \mathrm{Fix}(n), n \geq 1 \right\}$$

be the set of **periodic point barycentres**.

Lemma 3·1. *Ω is a compact convex subset of the unit disc in $\mathbf{C}$, and is symmetric about the real axis. Ω_{pp} is dense in Ω.*

Proof. The set

$$\left\{ \int_{S^1} z \, d\mu(z) : \mu \text{ is a Borel probability measure} \right\}$$

is precisely the unit disc. Thus Ω is a subset of the unit disc.

The convexity of Ω follows from the convexity of $\mathcal{M}$, since the map $\mu \mapsto \int_{S^1} z \, d\mu(z)$ is affine. The compactness of Ω follows from the compactness of $\mathcal{M}$, since $\mu \mapsto \int_{S^1} z \, d\mu(z)$ is continuous. Ω_{pp} is dense in Ω because $\mathcal{M}_{pp}$ is dense in $\mathcal{M}$, again by the continuity of $\mu \mapsto \int_{S^1} z \, d\mu(z)$.

If $\mu \in \mathcal{M}$ then we can define the probability measure $\bar{\mu}$ by $\bar{\mu}(A) = \mu(\bar{A})$, where $\bar{A}$ denotes the set of complex conjugates of the points in A. Note that $\bar{\mu} \in \mathcal{M}$. If $w = \int_{S^1} z \, d\mu(z)$, then its complex conjugate $\bar{w} \in \Omega$, since $\bar{w} = \int_{S^1} z \, d\bar{\mu}(z)$. Thus Ω is symmetric about the real axis. $\square$

Remark. Lemma 3·1 means that Ω will be completely determined by its **extremal points** $E(\Omega)$. We will conjecture (see Corollary 3·29) that every point on the topological boundary $\partial\Omega$ is an extremal point, so that $\partial\Omega$ contains no line segments. The density of the periodic point barycentres means that our first approach to the study of Ω will be via Ω_{pp}. The symmetry of Ω allows us to restrict attention to those barycentres lying in the (closed) upper half plane.

Note that $1 \in \mathbf{C}$ is an extremal point of Ω. It is the barycentre of the Dirac measure concentrated on the fixed point $1 \in S^1$ (corresponding to the fixed point $0 \in K$).

Definition 3·3. *For any $w \in \Omega$ we define the convex set*

$$\mathcal{M}(w) = \{\mu \in \mathcal{M} : \int_{S^1} z \, d\mu(z) = w\}.$$

We also define

$$\mathcal{M}(\partial\Omega) = \bigcup_{w \in \partial\Omega} \mathcal{M}(w), \quad \mathcal{M}(E(\Omega)) = \bigcup_{w \in E(\Omega)} \mathcal{M}(w) .$$

Lemma 3·2. *Suppose $\mathcal{M}(w)$ is a singleton set containing an ergodic measure. Then $w \in E(\Omega)$.*

Proof. Suppose $w \notin E(\Omega)$. Then we can find $\mu_1, \mu_2 \in \mathcal{M}$ with $\mu_1 \neq \mu_2$, and $\alpha \in (0,1)$, such that

$$w = \alpha \int_{S^1} z \, d\mu_1(z) + (1-\alpha) \int_{S^1} z \, d\mu_2(z) .$$

This implies that the non-ergodic measure $\alpha\mu_1 + (1-\alpha)\mu_2$ belongs to $\mathcal{M}(w)$, a contradiction. □

Lemma 3·3. *If $w \in E(\Omega)$ then $\mathcal{M}(w)$ contains an ergodic measure.*

Proof. Since $\mathcal{M}(w)$ is convex, we can choose μ to be an extremal point of $\mathcal{M}(w)$. We will show that in fact μ is extremal in $\mathcal{M}$, and therefore ergodic.

Suppose, for a contradiction, that we can find distinct measures $\mu_1, \mu_2 \in \mathcal{M}$ and $\alpha \in (0,1)$ such that

$$\mu = \alpha\mu_1 + (1-\alpha)\mu_2 . \tag{3·1}$$

Write $w_1 = \int_{S^1} z \, d\mu_1(z)$ and $w_2 = \int_{S^1} z \, d\mu_2(z)$. If $w_1 \neq w_2$, then integrating (3·1) around S^1 gives $w = \alpha w_1 + (1-\alpha)w_2$, which contradicts the fact that w is extremal in Ω.

So we must have $w_1 = w_2 = w'$, say. Integrating (3·1) around S^1 gives $w = \alpha w' + (1-\alpha)w' = w'$. Thus μ_1, μ_2 are distinct measures in $\mathcal{M}(w)$. But then (3·1) gives a contradiction, since μ is extremal in $\mathcal{M}(w)$. □

Remark. If w is not extremal in Ω then an extremal measure in $\mathcal{M}(w)$ need not be extremal in $\mathcal{M}$. However, in Corollary 3·28 we show that if w is an interior point of Ω, then $\mathcal{M}(w)$ does contain an ergodic measure.

Section 3·3. Maximal Measures and Strictly Maximal Orbits.

In §3·9 we will study $\partial\Omega$ via a family f_θ of real-valued analytic functions of the circle. In this section we develop the relevant concepts, but in the more general context of Hölder continuous functions. We let $\mathcal{H} = \mathcal{H}(K, \mathbf{R}) = \mathcal{H}^\alpha(K, \mathbf{R})$ denote the space of all real-valued α-Hölder functions, equipped with the usual Banach norm $\|\cdot\| = \|\cdot\|_\alpha + \|\cdot\|_\infty$. Here

$$\|f\|_\alpha = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x,y)^\alpha} ,$$

while $\|f\|_\infty = \sup_{x \in K} |f(x)|$ is the uniform norm.

We usually work with the Hölder norm $\|\cdot\|$. However, since several of the results in this chapter are valid for all continuous functions, we sometimes use the uniform norm.

Given a function $f \in \mathcal{H}(K, \mathbf{R})$, we write

$$f^n(x) = \sum_{i=0}^{n-1} f(T^i x).$$

Definition 3·4. *Define*

$$\Omega(f) = \left\{ \int f d\mu : \mu \in \mathcal{M} \right\}.$$

Since $\mu \mapsto \int f \, d\mu$ is continuous and affine, and $\mathcal{M}$ is compact and convex, then $\Omega(f)$ is also compact and convex. That is, $\Omega(f)$ is a closed interval. Clearly $\Omega(f)$ is contained in the closed interval $[\inf f, \sup f]$.

Definition 3·5. *If $u \in L^1(K, \mathbf{R})$ and $a \in \mathbf{R}$, then we say the function $u \cdot T - u + a$ is an* **essential coboundary**. *We say two functions $f, g \in \mathcal{H}$ are* **essentially cohomologous** *if they differ by an essential coboundary.*

If $f = u \cdot T - u + a$ is an essential coboundary then $\int f \, d\mu = a$ for *all* $\mu \in \mathcal{M}$, so that $\Omega(f)$ is the singleton $\{a\}$. Since $f \in \mathcal{H}$, then by Livsic's Theorem (see Livsic [31]), this trivial situation *only* arises when f is an essential coboundary.

We remark that two recent papers (Blokh [6], Ziemian [75]) investigate the set $\Omega(f)$ in a much more general context. Here the function f can be vector-valued, and typically satisfies weaker regularity properties. There are certain similarities with our work, though the results contained in these papers are not sharp enough to have any bearing on our problem.

A more fruitful approach is to use thermodynamic formalism to study $\Omega(f)$ further. This approach was pursued by Coelho [12]. First we remind the reader of some thermodynamic concepts.

Definition 3·6. *Given $g \in \mathcal{H}(K, \mathbf{R})$ we define its* **pressure** *$P(g)$ by*

$$P(g) = \sup_{\mu \in \mathcal{M}} \left(h(\mu) + \int g \; d\mu \right),$$

where $h(\mu)$ is the entropy of T with respect to μ.

An **equilibrium state** *for g is a measure $m_g \in \mathcal{M}$ satisfying*

$$h(m_g) + \int g \; dm_g = P(g).$$

We remark that since g is Hölder then it has a unique equilibrium state. The equilibrium state is ergodic, fully supported, and has positive entropy. Further details can be found in Walters [72].

Lemma 3·4. *The map $\mathcal{H} \to \mathcal{M}$, which takes each function f to its unique equilibrium state m_f, is continuous (with respect to the Hölder topology on $\mathcal{H}$, and the weak* topology on $\mathcal{M}$).*

Proof. This uses the theory of Ruelle-Perron-Frobenius operators, as introduced in §1·2. We can define such an operator in terms of the Hölder function f. Then m_f is the eigenvector corresponding to the unique maximal eigenvalue λ_f of the adjoint operator (for details see Parry & Pollicott [50]). Since this eigenvalue is isolated and simple, perturbation theory tells us that both λ_f and m_f vary continuously with f. $\square$

For a fixed $f \in \mathcal{H}$, the map $t \mapsto P(tf)$ is an analytic function of the real variable t. Its derivative is given (see Parry & Pollicott [50]) by the formula

$$\frac{d}{dt} P(tf)|_{t=t_0} = \int f \; dm_{t_0 f} \; . \tag{3·2}$$

We have the following characterisation of $\Omega(f)$.

Proposition 3·5. *(Bohr & Rand, [7]) If $f \in \mathcal{H}(K, \mathbf{R})$, then*

$$int(\Omega(f)) = \left\{ \frac{d}{dt} P(tf)|_{t=t_0} : t_0 \in \mathbf{R} \right\}.$$

Proof. The result was first proved by Bohr & Rand [7], though the notation in Proposition III.3.2 of Coelho [12] is closer to ours. □

Definition 3·7. *If $b \in \Omega(f)$, then we define*

$$\mathcal{M}(f, b) = \left\{ \mu \in \mathcal{M} : \int f \, d\mu = b \right\}.$$

Proposition 3·6. *(Lanford, [29]) Suppose $f \in \mathcal{H}$ is not an essential coboundary, and $b \in int(\Omega(f))$. Then*

(a) the equilibrium state $m_{t_0 f}$ belongs to $\mathcal{M}(f, b)$, where $t_0 \in \mathbf{R}$ is the unique value satisfying $b = \frac{d}{dt} P(tf)|_{t=t_0} = \int f \, dm_{t_0 f}$,

(b) $m = m_{t_0 f}$ is the unique measure in $\mathcal{M}(f, b)$ satisfying $h(m) = \sup_{\mu \in \mathcal{M}(f,b)} h(\mu)$,

(c) $h(m_{t_0 f}) = P(t_0(f - b))$.

Proof. This was first proved by Lanford [29]. See also Proposition III.1.1 of Coelho [12]. □

The following positive functional has some similarities with pressure.

Definition 3·8. *Define $Q : \mathcal{H}(K, \mathbf{R}) \to \mathbf{R}$ by*

$$Q(f) = \sup_{\mu \in \mathcal{M}} \int f \, d\mu = \sup \ \Omega(f).$$

We immediately have

Proposition 3·7. *If $f \in \mathcal{H}(K, \mathbf{R})$ then*

$$Q(f) = \lim_{t_0 \to \infty} \int f \, dm_{t_0 f} = \lim_{t_0 \to \infty} \frac{d}{dt} P(tf)|_{t=t_0} .$$

Proof. The proof is the same as that used for Proposition 3·5. The second equality is just equation (3·2). □

Lemma 3·8. *$Q : \mathcal{H}(K, \mathbf{R}) \to \mathbf{R}$ is Lipschitz with respect to the uniform norm, with Lipschitz constant 1.*

Proof. For $f, g \in \mathcal{H}(K, \mathbf{R})$ we have

$$\begin{aligned} |Q(f) - Q(g)| &= \left| \sup_{\mu \in \mathcal{M}} \int f \, d\mu - \sup_{\mu \in \mathcal{M}} \int g \, d\mu \right| \\ &\le \sup_{\mu \in \mathcal{M}} \left| \int f \, d\mu - \int g \, d\mu \right| \\ &\le \sup_{\mu \in \mathcal{M}} \int |f - g| \, d\mu \\ &\le \|f - g\|_\infty . \quad \square \end{aligned}$$

Definition 3·9. *For $f \in \mathcal{H}(K, \mathbf{R})$, a measure $m \in \mathcal{M}$ is called* f-**maximal** *if $\int f dm = Q(f)$. We let $\mathcal{M}(f)$ denote the set of f-maximal measures. With the notation of Definition 3·7 we have $\mathcal{M}(f) = \mathcal{M}(f, Q(f))$.*

Note that $\mathcal{M}(f)$ is always a non-empty compact convex set. It is straightforward to prove (though we will not use the fact) that $\mathcal{M}(f)$ is equal to the set of tangent functionals to Q at f. This terminology is explained on page 224 of Walters [72], and the analogous result is proved for pressure. A consequence of this (see page 226 of Walters [72], or page 450 of Dunford & Schwartz [15]) is the existence of a dense subset $\mathcal{H}' \subset \mathcal{H}$ such that $\mathcal{M}(f)$ is a singleton set for all $f \in \mathcal{H}'$. In §3·9 we introduce a family f_θ of trigonometric functions, and in §3·10 we conjecture that $\mathcal{M}(f_\theta)$ is always a singleton set.

Note that if $x \in \mathrm{Fix}(n)$ then

$$\int f \, d\mu_x = \frac{1}{n} f^n(x).$$

By analogy with Lemma 3·1, we note that periodic point integrals are dense in $\Omega(f)$.

A periodic orbit is the simplest example of an orbit whose closure supports a *unique* T-invariant Borel probability measure (i.e. the restriction of T to $\overline{\mathcal{O}(x)}$ is uniquely ergodic). In later sections we will be interested in more general orbits $\mathcal{O}(x)$ with this property. We denote the unique measure by $\mu_{\mathcal{O}(x)}$. By Theorem 6.19 of Walters [72], if the restriction of T to $\overline{\mathcal{O}(x)}$ *is* uniquely ergodic, then the ergodic average $\lim_{n\to\infty} \frac{1}{n} f^n(x)$ exists, and equals $\int f \, d\mu_{\mathcal{O}(x)}$.

Definition 3·10. *Let $x \in K$, and suppose that the restriction of T to $\overline{\mathcal{O}(x)}$ is uniquely ergodic. We say the orbit $\mathcal{O}(x)$ is* **strictly maximal** *(for f) if for all $k \geq 1$, and for all period-k points $z \notin \mathcal{O}(x)$, we have*

$$\lim_{n\to\infty} \frac{1}{n} f^n(x) > \frac{1}{k} f^k(z). \tag{3·3}$$

Lemma 3·9. *Let $f \in \mathcal{H}$ and $x \in K$, and suppose that the restriction of T to $\overline{\mathcal{O}(x)}$ is uniquely ergodic. Then*

(a) If the orbit $\mathcal{O}(x)$ is strictly maximal for f, then $\mu_{\mathcal{O}(x)}$ is a maximal measure for f.

(b) If $\mu_{\mathcal{O}(x)}$ is the unique maximal measure for f, then the orbit $\mathcal{O}(x)$ is strictly maximal for f.

Proof. (a) Suppose $\mu_{\mathcal{O}(x)}$ is not a maximal measure. So $\int f \, d\mu_{\mathcal{O}(x)} < Q(f)$. Since periodic point integrals are dense in $\Omega(f)$, we can find a period-k point $z \notin \mathcal{O}(x)$ such that

$$\int f \, d\mu_{\mathcal{O}(x)} < \frac{1}{k} f^k(z) < Q(f).$$

By the ergodic theorem this means that

$$\lim_{n\to\infty} \frac{1}{n} f^n(x) < \frac{1}{k} f^k(z) ,$$

contradicting the fact that $\mathcal{O}(x)$ is strictly maximal.

(b) Suppose $\mathcal{O}(x)$ is not strictly maximal. Then there exists a period-k point $z \notin \mathcal{O}(x)$ such that

$$\lim_{n\to\infty} \frac{1}{n} f^n(x) \leq \frac{1}{k} f^k(z) ,$$

By the ergodic theorem this means that

$$\int f \, d\mu_{\mathcal{O}(x)} \leq \int f \, d\mu_z ,$$

contradicting the fact that $\mu_{\mathcal{O}(x)}$ is the unique maximal measure. □

Lemma 3·10. *Suppose $f \in \mathcal{H}$ and $x \in \mathrm{Fix}(n)$. Then $\mathcal{O}(x)$ is strictly maximal for f if and only if for all $k \geq 1$, and for all period-k points $z \notin \mathcal{O}(x)$, we have*

$$f^{kn}(x) > f^{kn}(z). \tag{3·4}$$

Proof. Since $x \in \mathrm{Fix}(n)$ then (3·3) is equivalent to

$$\frac{1}{n} f^n(x) > \frac{1}{k} f^k(z).$$

Multiplication by kn gives (3·4). $\square$

We now introduce a stronger concept for strictly maximal periodic orbits.

Definition 3·11. *Suppose $f \in \mathcal{H}$ and $x \in \mathrm{Fix}(n)$. We say the periodic orbit $\mathcal{O}(x)$ is* **uniformly strictly maximal** *(for f) if there exists $c > 0$ such that for all $k \geq 1$, and for all period-k points $z \notin \mathcal{O}(x)$, we have*

$$f^{kn}(x) - f^{kn}(z) > c. \tag{3·5}$$

We say a point $x \in \mathrm{Fix}(n)$ is **(uniformly) strictly maximal** *if $\mathcal{O}(x)$ is a (uniformly) strictly maximal orbit.*

We remark that (3·5) is equivalent to

$$\frac{1}{n} f^n(x) - \frac{1}{k} f^k(z) > \frac{c}{nk}\ .$$

An interpretation of this is that the uniformly strictly maximal periodic point integral is *badly approximable* by other periodic point integrals. The smaller the period n is, the harder it is to approximate $\frac{1}{n} f^n(x)$. In §3·9 we use a family f_θ of trigonometric functions to study the boundary of Ω, and in §3·10 we conjecture that almost every f_θ has a uniformly strictly maximal periodic orbit. The difficulty in approximating a uniformly strictly maximal periodic point integral is seen in Appendix D, where the points $1, -\frac{1}{2} \in \mathbf{C}$ (which correspond to orbits of period 1 and 2, respectively) appear somewhat isolated.

Note that if the periodic orbit $\mathcal{O}(x)$ is uniformly strictly maximal for some $f \in \mathcal{H}$, then $\mathcal{O}(x)$ is also uniformly strictly maximal for any $f' \in \mathcal{H}$ sufficiently close to f in the Hölder topology. The proof of this uses a standard application of the closing lemma (see Lemma 3·12), and is identical to part of the proof of Proposition 3·13.

Definition 3·12. *Let $\mu \in \mathcal{M}$. We say $f \in \mathcal{H}(K, \mathbf{R})$ is* **recurrent** *with respect to μ if for all $\epsilon > 0$, and all Borel sets B with $\mu(B) > 0$, there exists $k \geq 1$ such that*

$$\mu(B \cap T^{-k}B \cap \{y : |f^k(y)| < \epsilon\}) > 0.$$

We will need the following theorem, which is due to G. Atkinson.

Theorem 3·11. *(Atkinson, [4]) Let $\mu \in \mathcal{M}$ be non-atomic, and let $f \in \mathcal{H}(K, \mathbf{R})$. Then f is recurrent with respect to μ if and only if $\int f \, d\mu = 0$.* □

Recall that a *dyadic interval* of K is an interval of the form $[r/2^n, (r+1)/2^n)$, for some $n \geq 1$ and $0 \leq r \leq 2^n - 1$. The following lemma is a standard result of hyperbolic dynamics.

Lemma 3·12. *(Closing Lemma)*

Let $I \subset K$ be a dyadic interval of length $\beta > 0$. Suppose $y, T^k y \in I$. Then there exists $z \in \mathrm{Fix}(T^k)$ such that

$$d(T^i z, T^i y) \leq (1/2)^{k-i}\beta \qquad \textit{for all} \quad i = 0, 1 \ldots k-1.$$

Proof. See page 269 of Katok & Hasselblatt [25], for example. □

The following result is a strengthening of Lemma 3·9.

Proposition 3·13. *Suppose $f \in \mathcal{H}(K, \mathbf{R})$ has a uniformly strictly maximal period-n orbit $\mathcal{O}(x)$. Then the corresponding periodic point measure μ_x is the unique maximal measure for f.*

Proof. The measure μ_x is certainly maximal, by Lemma 3·9 (a). It remains to prove uniqueness.

Suppose, for a contradiction, that μ is some other maximal measure. Certainly μ cannot be purely atomic, for such measures are concentrated on (the countable union of) periodic orbits. The integral of f with respect to any such measure will be less than the integral of f with respect to μ_x (since $\mathcal{O}(x)$ is the strictly maximal periodic orbit), and therefore not a maximal measure.

In fact μ cannot be atomic, for then it could be written as a convex combination of a purely atomic measure μ_1 and a non-atomic measure μ_2. If μ were maximal then so would μ_1 be, contradicting the above.

It remains to eliminate the possibility that μ is a non-atomic maximal measure, so let us suppose that μ is non-atomic.

By replacing f with $f - \int f d\mu_x$ if necessary, we may assume that $f^n(x) = 0$. Thus we have $\int f d\mu_x = 0$. The uniform strict maximality of x means that for any period-k point z (not in the orbit of x) we have $f^{kn}(z) < -c < 0$ (where $c > 0$ is the uniform constant in Definition 3·11).

For any $0 < \delta < 1/2n$, let N_δ be the (proper) subset of K obtained by putting an interval of radius δ around each point in $\mathcal{O}(x)$. That is,

$$N_\delta = \bigcup_{r=0}^{n-1} (T^r x - \delta, T^r x + \delta).$$

Suppose f has Hölder exponent $\alpha \in (0,1]$ and corresponding Hölder constant $M > 0$.

$$\text{Choose} \quad 0 < \epsilon < c/2n \qquad \text{and} \qquad 0 < \beta < \min\left(2\delta \ , \ 2\left(\frac{\epsilon}{2M}\right)^{1/\alpha}\right).$$

Let $B \subset K \setminus N_\delta$ be an interval contained in a dyadic interval of length less than β. Let us assume, for a contradiction, that $\mu(B) > 0$.

Since μ is non-atomic and $\int f d\mu = 0$, Theorem 3·11 tells us that f is recurrent with respect to μ. Therefore there exists $k > 0$ such that

$$\mu(B \cap T^{-k}B \cap \{y : |f^k(y)| < \epsilon\}) > 0.$$

In particular, there exists $y \in B$ such that $T^k y \in B$ and $|f^k(y)| < \epsilon$.

By Lemma 3·12 we know there exists a period-k point z such that

$$d(T^i z, T^i y) \leq (1/2)^{k-i}\beta \qquad \text{for } i = 0, 1, \ldots, k-1. \tag{3·6}$$

In particular, for $i = 0$ we have $d(z, y) \leq \beta/2$.

Thus

$$\begin{aligned} d(z, \mathcal{O}(x)) &\geq d(y, \mathcal{O}(x)) - d(z, y) \\ &> \delta - \beta/2 \\ &> 0 \quad \text{since } \beta < 2\delta. \end{aligned}$$

Therefore the period-k point z does not lie in the orbit of x.

Now we want to show that $-2\epsilon < f^k(z) < 0$. By (3·6) we see that, for $i = 0, 1, \ldots, k-1$,

$$|f(T^i z) - f(T^i y)| \;\leq\; M\left[(1/2)^{k-i}\beta\right]^{\alpha} \;=\; M\left[(1/2)^{\alpha}\right]^{k-i}\beta^{\alpha}.$$

Therefore we have

$$\begin{aligned} |f^k(z) - f^k(y)| &\leq \sum_{i=0}^{k-1} |f(T^i z) - f(T^i y)| \\ &\leq M\beta^{\alpha} \sum_{i=0}^{k-1} \left[\left(\frac{1}{2}\right)^{\alpha}\right]^{k-i} \\ &< M\beta^{\alpha} \frac{\left(\frac{1}{2}\right)^{\alpha}}{1 - \left(\frac{1}{2}\right)^{\alpha}} \\ &< 2M\left(\frac{\beta}{2}\right)^{\alpha} \\ &< \epsilon \quad \text{since } \beta < 2\left(\frac{\epsilon}{2M}\right)^{1/\alpha}. \end{aligned}$$

It follows that

$$|f^k(z)| \leq |f^k(z) - f^k(y)| + |f^k(y)| < \epsilon + \epsilon = 2\epsilon.$$

Moreover, since z is periodic and not on the strictly maximal orbit then $f^k(z)$ must be negative. So we have

$$-2\epsilon < f^k(z) < 0.$$

Thus

$$-2n\epsilon < nf^k(z) = f^{kn}(z) < 0.$$

Since $\epsilon < c/2n$ we obtain

$$-c < f^{kn}(z) < 0,$$

which contradicts the uniform strict maximality of the period-n point x.

So our assumption that $\mu(B) > 0$ led to a contradiction. Therefore any interval $B \subset K \setminus N_\delta$ contained in a dyadic interval of length less than β has zero μ-measure. Writing $K \setminus N_\delta$ as a union of such intervals we see that $\mu(K \setminus N_\delta) = 0$. Letting $\delta \to 0$ we see that $\mu(K \setminus \mathcal{O}(x)) = 0$. So μ is concentrated on $\mathcal{O}(x)$, contradicting our assumption that μ is non-atomic. This completes the proof. □

A weaker hypothesis on f gives the following weaker conclusion. This result was proved by Coelho [12].

Proposition 3·14. *(Coelho, [12]) Suppose $f \in \mathcal{H}(K, \mathbf{R})$ is not an essential coboundary. Any maximal measure for f cannot be fully supported.*

Proof. This is an immediate consequence of Proposition III.3.4 in Coelho [12]. □

Section 3·4. Empirical Results.

By Lemma 3·1 we know that Ω_{pp} is dense in Ω. Therefore our first approach to the study of Ω is to calculate periodic point barycentres, for low periods.

The fixed point $0 \in K$ corresponds to Dirac measure concentrated on $1 \in S^1$. The barycentre of this measure is the point $1 \in \mathbf{C}$.

The period-2 orbit $\{1/3,\ 2/3\}$ has barycentre

$$\frac{1}{2}\left(e^{2\pi i/3} + e^{4\pi i/3}\right) = -\frac{1}{2}.$$

There are two orbits of least period 3. The barycentre of $\{1/7,\ 2/7,\ 4/7\}$ is

$$\frac{1}{3}\left(e^{2\pi i/7} + e^{4\pi i/7} + e^{8\pi i/7}\right) = -\frac{1}{6} + \frac{\sqrt{7}}{6}\, i \simeq -0.1666666 + 0.4409585\, i\ ,$$

while the barycentre of the conjugate orbit $\{6/7,\ 5/7,\ 3/7\}$ is the complex conjugate

$$-\frac{1}{6} - \frac{\sqrt{7}}{6}\, i \simeq -0.1666666 - 0.4409585\, i\ .$$

Let Ω_3 be the convex hull of the four barycentres we have calculated.

More generally, let Ω_j denote the convex hull of those periodic point barycentres of period less than or equal to j. Then each Ω_j is a convex set contained in Ω. Moreover, the Ω_j form an increasing sequence whose union is dense in Ω. The boundary of this union is precisely $\partial\Omega$.

Let E_j denote the set of extremal points of Ω_j. Experimental evidence (see Appendix C) suggests that the E_j also form an increasing sequence. That is, if w is an extremal point of some Ω_k, then it is an extremal point of all Ω_j, $j \geq k$. It would follow that w is an extremal point of Ω. Therefore we believe that the sets E_j give information about the extremal points of Ω, and are worth studying.

It will be convenient to introduce symbolic codes for periodic orbits. Every $x \in K$ can be written

$$x = \sum_{j=1}^{\infty} \frac{x_j}{2^j} \quad , \quad x_j \in \{0, 1\}.$$

The sequence $\underline{x} = (x_1, x_2, \ldots)$ is called the **symbolic code** for x. If we forbid those sequences ending in an infinite string of zeros, then each $x \in K$ has a unique symbolic code. If now $x \in \mathrm{Fix}(n)$ then its symbolic code is given by repeating the length-n block $x_1 \ldots x_n$. In this case we refer to $x_1 \ldots x_n$ as the symbolic code for x, and write $\underline{x} = x_1 \ldots x_n$. Moreover, the symbolic code of any iterate $T^j x \in \mathcal{O}(x)$ is $x_{j+1} \ldots x_n x_1 \ldots x_j$, a cyclic shift of the symbolic code for x. We define the symbolic code of the period-n orbit $\mathcal{O}(x)$ to be the symbolic code of the *smallest* point (considered as an element of $[0, 1)$) on the orbit. For example, the period-5 orbit $\{5/31,\ 10/31,\ 20/31,\ 9/31,\ 18/31\}$ has symbolic code 00101.

We have seen that the barycentres of all the periodic orbits of period less than or equal to 3 are extremal points of Ω_3. The following table shows that this is not the case for orbits of least period 4.

Orbit	Symbolic code	Barycentre
$\mathcal{O}(1/15)$	0001	$0.125 + 0.4841229\, i$
$\mathcal{O}(7/15)$	0111	$0.125 - 0.4841229\, i$
$\mathcal{O}(1/5)$	0011	-0.25

Note that the barycentres of the first two orbits *are* extremal points of Ω_4. Since the orbits are conjugate, their barycentres are complex conjugates. The barycentre of the orbit $\{1/5,\ 2/5,\ 4/5,\ 3/5\}$ is not an extremal point of Ω_4.

Henceforth we will exploit the symmetry of Ω about the real axis, and only list those periodic orbits whose integrals have non-negative imaginary part. For such orbits of least period 5 we have the following table.

Symbolic code	Barycentre
00001	$0.3083872 + 0.443599\, i$
00011	$-0.0786801 + 0.1745122\, i$
00101	$-0.329707 + 0.2876896\, i$

Note that the barycentres of 00001 and 00101 are extremal points of Ω_5, but the same is not true of 00011.

The orbits of least period 6 with non-negative imaginary part are

Symbolic code	Barycentre
000001	$0.4285381 + 0.3919765\, i$
000011	$0.0833333 + 0.2204793\, i$
000101	$-0.124271 + 0.3709144\, i$
000111	0
001101	$-0.3042671 + 0.0210621\, i$

The only extremal point of Ω_6 in the above is given by 000001.

Using the computer program Mathematica we calculated all the extremal barycentres for periodic orbits up to period 19. In Appendix C we list those periodic orbits whose barycentres are extremal and have non-negative imaginary part (the symmetry of Ω also gives us those with negative imaginary part). There are 120 extremal points of Ω_{19}, two of which lie on the real line.

For each $n \leq 19$ the program verified that the extremal points of Ω_{n-1} were also extremal points of Ω_n. The program involved:

1. Computation of all barycentres of periodic orbits of a given period $n \leq 19$.
2. Checking which of these barycentres lay outside the convex region Ω_{n-1}.
3. Checking that the extremal points of Ω_{n-1} were also extremal points of Ω_n.

It was observed that many triples of extremal points are 'almost collinear'. This phenomenon can be seen in Appendix D, where we plot the extremal points of Ω_{19}. The calculations used a precision of twelve decimal places, and this was sufficient to verify that the 'almost collinear' points are not in fact collinear. In §3·6 we describe a different way of approximating Ω. The extremal points generated by this method, even in the early stages, are even closer to being collinear. We need to continually improve the precision of the calculations to check that the points are not in fact collinear.

Since we believe that all extremal points of Ω_{19} are actually extremal points of Ω, the plot in Appendix D seems to suggest that $\partial\Omega$ is non-differentiable. At the points 1 and $-\frac{1}{2}$ the non-differentiability seems particularly pronounced. Corollary 3·33, which relies on an unproved conjecture, states that $\partial\Omega$ contains a countable dense set of points of non-differentiability.

Section 3·5. Barycentres at the Origin.

We remarked in §3·2 that the barycentre of Lebesgue measure is the origin in $\mathbb{C}$. In §3·4 we saw that the periodic point measure concentrated on the period-6 orbit $\mathcal{O}(\frac{1}{9})$ (with symbolic code 000111) also has barycentre at $0 \in \mathbb{C}$. Thus the barycentre map $\mu \mapsto \int_{S^1} z \, d\mu(z)$ is certainly not injective. If we restrict this map to the countable dense subset $\mathcal{M}_{pp} \subset \mathcal{M}$, then it is still not injective, as the following construction will show. In fact we can find infinitely many periodic point measures whose barycentre is zero.

Suppose $x_1, x_2, x_3 \in K$ are three evenly spaced points on the circle (i.e. distance $\frac{1}{3}$ from each other). Then symmetry gives us

$$e^{2\pi i x_1} + e^{2\pi i x_2} + e^{2\pi i x_3} = 0.$$

The reason $\mathcal{O}(\frac{1}{9})$ gives a zero barycentre is that it decomposes into two sets of three evenly spaced points, namely $\{\frac{1}{9}, \frac{4}{9}, \frac{7}{9}\}$ and $\{\frac{2}{9}, \frac{5}{9}, \frac{8}{9}\}$.

In general if $x = \frac{1}{3(2^{2n}-1)}$ then the period-$6n$ orbit $\mathcal{O}(x)$ decomposes into $2n$ sets of three evenly spaced points. Thus the periodic point measure concentrated on $\mathcal{O}(x)$ gives zero barycentre.

Slightly more generally, suppose $3(2^{2n}-1)$ has a prime factor p with $p \equiv 1 \pmod 3$. If $x = \frac{p}{3(2^{2n}-1)}$, then again the period-$6n$ orbit $\mathcal{O}(x)$ decomposes into $2n$ sets of three evenly spaced points, and the corresponding barycentre is zero.

In the table below we list those orbits up to period 42 which correspond to zero barycentres by the above reasoning.

Period	Zero barycentre orbits		
6	$\mathcal{O}(1/9)$		
12	$\mathcal{O}(1/45)$		
18	$\mathcal{O}(1/189)$,	$\mathcal{O}(1/27)$	
24	$\mathcal{O}(1/765)$		
30	$\mathcal{O}(1/3069)$,	$\mathcal{O}(1/99)$	
36	$\mathcal{O}(1/12285)$,	$\mathcal{O}(1/1755)$,	$\mathcal{O}(1/945)$
42	$\mathcal{O}(1/49149)$,	$\mathcal{O}(1/1143)$,	$\mathcal{O}(1/387)$

Section 3·6. Ordered Orbits.

In §3·4 we introduced the convex approximations Ω_j to our set Ω. In Appendix C we list the extremal points of Ω_{19}, together with the symbolic codes of their corresponding periodic orbits. We remark that these symbolic codes all satisfy the following definition.

Definition 3·13. *A symbolic code* $(x_1, x_2, \ldots) \in \{0,1\}^{\mathbb{N}}$ *is said to be* **Sturmian** *if the number of 1's in any two sub-blocks of the same length differs by at most one.*

This Sturmian condition was first studied by Hedlund & Morse [41].

In fact, for all symbolic codes listed in Appendix C, if the number of 1's in two sub-blocks of the same length *does* differ, then it is the right-hand sub-block which has the extra 1. However, this is simply because the codes listed only correspond to extremal points with *non-negative* imaginary part. By swapping 0's and 1's we obtain the codes for those extremal points with non-positive imaginary part. These codes still satisfy the Sturmian condition, but now if there is a difference of 1's then it is the left-hand sub-block which has the extra 1.

There are many alternative characterisations of the Sturmian condition, which we list in the forthcoming Lemma 3·15. Before stating this lemma we describe a constructive method for generating all Sturmian codes. This process is reminiscent of the well-known method of constructing Farey fractions (see Hardy & Wright [21], for example), and involves the concatenation of finite codes.

Suppose $x \in \mathrm{Fix}(m)$ and $y \in \mathrm{Fix}(n)$, where $x < y$ (thinking of $x, y \in [0,1)$). Denote the corresponding symbolic codes by $\underline{x} = x_1 \ldots x_m$ and $\underline{y} = y_1 \ldots y_n$. We define the concatenation $\underline{x} \oplus \underline{y}$ by

$$\underline{x} \oplus \underline{y} = x_1 \ldots x_m y_1 \ldots y_n$$

If $x = \frac{a}{2^m - 1}$ and $y = \frac{b}{2^n - 1}$, then we will also write

$$x \oplus y = \frac{2^n a + b}{2^{m+n} - 1} \, .$$

It is a simple check that

$$\frac{a}{2^m - 1} < \frac{2^n a + b}{2^{m+n} - 1} < \frac{b}{2^n - 1} \, .$$

The **Farey tower** is the infinite tower consisting of countably many Farey levels indexed by the non-negative integers. Each Farey level consists of a finite sequence of symbolic codes, in lexicographic order (where $0 < 1$, and we read from left to right). We define the 0^{th} level to contain the codes $0 < 01$. If the j^{th} Farey level is

$$\underline{x}^1 < \underline{x}^2 < \underline{x}^3 < \ldots < \underline{x}^{r-1} < \underline{x}^r ,$$

then we define the $(j+1)^{th}$ level to be

$$\underline{x}^1 < \underline{x}^1 \oplus \underline{x}^2 < \underline{x}^2 < \underline{x}^2 \oplus \underline{x}^3 < \underline{x}^3 < \ldots < \underline{x}^{r-1} < \underline{x}^{r-1} \oplus \underline{x}^r < \underline{x}^r .$$

The first four levels of the Farey tower are as follows.

Level									
0	0	01							
1	0	001	01						
2	0	0001	001	00101	01				
3	0	00001	0001	0001001	001	00100101	00101	0010101	01

Let the **extended Farey tower** consist of the Farey tower itself, together with those Farey codes 'at infinity', i.e. those codes obtained through infinitely many concatenations but which are not eventually periodic sequences.

Before stating Lemma 3·15, we introduce some terminology.

If $x \in K$ with symbolic code $(x_1, x_2, \ldots)$, we define

$$\rho_n(x) = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

By the ergodic theorem we know that $\rho(x) = \lim_{n\to\infty} \rho_n(x)$ exists for Lebesgue almost every $x \in K$, and if so we call this the **rotation number** of x. Note that $\rho(x)$ just represents the average number of 1's in the symbolic code for x. For certain points x, however, this is a rotation number in the usual sense of the word, since we can associate to it a certain degree one circle map (see the remarks towards the end of §3·7).

Lemma 3·15. *Let $x \in K$. The following are equivalent.*

(a) The symbolic code for x is Sturmian.

(b) The symbolic code for x belongs to the extended Farey tower.

(c) The orbit $\mathcal{O}(x)$ is minimal and contained in some semi-circle $[\delta, \delta + \frac{1}{2}] \subset K$.

(d) The orbit $\mathcal{O}(x)$ is minimal and is ordered. That is, if a, b, c are points on $\mathcal{O}(x)$, then their cyclic order around K is preserved by T.

(e) The rotation number $\rho(x) = \rho$ exists, and the convergence of the averages $\rho_n(x)$ to ρ is faster than for any other $y \notin \mathcal{O}(x)$ with $\rho(y) = \rho$.

Proof. This is something of a folklore result. The article by Bullett & Sentenac [10] gives an overview of the various equivalent definitions. The approach (e) was first studied by Veerman [70], [71]. □

Definition 3·14. *If $\mathcal{O}(x)$ satisfies any of the five equivalent conditions in Lemma 3·15 then we call it an* **ordered** *orbit.*

We remark (see Bullett & Sentenac [10]) that if $\mathcal{O}(x)$ is a non-periodic ordered orbit contained in the semi-circle $[\delta, \delta + \frac{1}{2}]$, then both of the endpoints δ, $\delta + \frac{1}{2}$ belong to the orbit. We can then define the symbolic code of $\mathcal{O}(x)$ to be the symbolic code of the smaller (as elements of $[0, 1)$) endpoint.

We will be interested in measures supported on the closure of an ordered orbit. We have the following results.

Lemma 3·16. *(Bullett & Sentenac, [10]) The closure of an ordered orbit has zero Hausdorff dimension.* □

Proposition 3·17. *If $\mathcal{O}(x)$ is an ordered orbit, then there is a unique T-invariant Borel probability measure supported on $\overline{\mathcal{O}(x)}$.*

Proof. Since $\mathcal{O}(x)$ is ordered, the corresponding symbolic code $\underline{x} = (x_1, x_2, \ldots)$ is Sturmian, by Lemma 3·15. So the number of 1's in any two sub-blocks of $\underline{x}$ of the same length differs by at most one. Similarly, if B is any given string of 0's and 1's, then the number of occurrences of B in any two sub-blocks of $\underline{x}$ of the same length also differs by at most one. By §10 of Oxtoby [45], this means that for all continuous functions f, there exists a constant $c \in \mathbf{R}$ such that

$$\sup_{i \geq 0} \left| \frac{1}{n} f^n(T^i x) - c \right| \to 0 \quad \text{as } n \to \infty.$$

Therefore the ergodic averages $\frac{1}{n} f^n$ converge *uniformly* on the space $\overline{\mathcal{O}(x)}$. By Theorem 6.19 of Walters [72], this means that the restriction of T to $\overline{\mathcal{O}(x)}$ is uniquely ergodic.

□

Corollary 3·18. *If $\mathcal{O}(x)$ is an ordered orbit, and if $\mu \in \mathcal{M}$ is the T-invariant Borel probability measure supported on $\overline{\mathcal{O}(x)}$, then μ is ergodic and has zero entropy.*

Proof. Let S denote the restriction of T to $\overline{\mathcal{O}(x)}$. Ergodicity of μ is clear, since it is the unique invariant probability measure of S.

By Proposition III.1 of Furstenberg [18] we have

$$HD(\overline{\mathcal{O}(x)}) = \frac{h_{top}(S)}{\log 2},$$

where $HD(\cdot)$ denotes Hausdorff dimension, and $h_{top}(\cdot)$ denotes topological entropy. Thus $h_{top}(S) = 0$, by Lemma 3·16. Then $h(\mu) = 0$ by the variational principle (see Theorem 8.6 of Walters [72]). □

It has been verified that all codes corresponding to extremal points of Ω_{19} appear in the Farey tower (though for those extremal points with negative imaginary part we must consider the 'conjugate' Farey tower obtained by swapping 0's and 1's).

Let $\mathcal{F}_j$ denote the set of symbolic codes on the j^{th} Farey level, and $\mathcal{B}(\mathcal{F}_j)$ the set of corresponding barycentres. Let $\Omega(\mathcal{F}_j)$ denote the convex hull of $\mathcal{B}(\mathcal{F}_j)$. Note that $\mathcal{F}_j$, $\mathcal{B}(\mathcal{F}_j)$, $\Omega(\mathcal{F}_j)$ are all increasing sequences of sets.

Using the Mathematica computer program we calculated $\mathcal{B}(\mathcal{F}_j)$ for all $j \leq 9$. We verified that each element of $\mathcal{B}(\mathcal{F}_j)$ is an extremal point of $\Omega(\mathcal{F}_j)$. This 'persistence' of extremal points leads us to conjecture that the extremal points of each $\Omega(\mathcal{F}_j)$ are in fact extremal points of Ω. In other words, that every Sturmian code corresponds to an element of $E(\Omega)$.

It was observed that if $\underline{x} < \underline{x} \oplus \underline{y} < \underline{y}$ are consecutive codes on some j^{th} Farey level, $j \leq 9$, then

$$\mathrm{Re} \int_{S^1} z \, d\mu_x(z) \; > \; \mathrm{Re} \int_{S^1} z \, d\mu_{x \oplus y}(z) \; > \; \mathrm{Re} \int_{S^1} z \, d\mu_y(z) \, .$$

It was observed that often three extremal points of $\Omega(\mathcal{F}_j)$ are very close to being collinear (if they *were* collinear then the point in the middle would not be extremal). This phenomenon was discussed, in the context of the approximations Ω_j, in §3·4, though in the Farey construction it is more pronounced, and occurs at an earlier stage. For example, an accuracy of 12 decimal places is sufficient to verify that no three points of $\mathcal{B}(\mathcal{F}_6)$ are collinear. However, the precision must be improved to check the same fact for $\mathcal{B}(\mathcal{F}_7)$. Increasing the accuracy to 40 decimal places verifies that no three points of $\mathcal{B}(\mathcal{F}_8)$ are collinear, but this precision fails for $\mathcal{B}(\mathcal{F}_9)$. We believe however, that for all $j \geq 0$, no three points of $\Omega(\mathcal{F}_j)$ are collinear.

Section 3·7. The Devil's Staircase.

For $\lambda \in K$, let $C_\lambda = [\lambda - \frac{1}{4}, \lambda + \frac{1}{4}] \subset K$ be the closed semi-circle centred around λ. Bullett & Sentenac [10], following earlier work of Gambaudo et al. [19] and Veerman [70], [71], proved the following result.

Proposition 3·19. *(Bullett & Sentenac, [10]) Each semi-circle $C_\lambda \subset K$ contains a unique minimal closed T-invariant set A_λ.* $\square$

By Lemma 3·15, each A_λ is just the closure of some ordered orbit. We can think of $\lambda \mapsto A_\lambda$ as a map, and refer to the sets A_λ as the *values* of the map. The following result is strongly analogous to Conjecture II in §3·10.

Proposition 3·20. *(Bullett & Sentenac, [10]) There is a sequence of disjoint non-trivial intervals $[\lambda_i^-, \lambda_i^+] \subset K$ such that*

(a) $\cup_{i=1}^{\infty}(\lambda_i^-, \lambda_i^+)$ has full Lebesgue measure. Its complement is a Cantor set of zero Hausdorff dimension.

(b) There is a one-to-one correspondence between intervals in the sequence and periodic ordered orbits. If $[\lambda_i^-, \lambda_i^+]$ is an interval, and $\mathcal{O}(x)$ is the corresponding periodic ordered orbit, then $A_\lambda = \mathcal{O}(x)$ for all $\lambda \in [\lambda_i^-, \lambda_i^+]$.

(c) There is a one-to-one correspondence between points of $\left[\cup_{i=1}^{\infty}[\lambda_i^-, \lambda_i^+]\right]^c$ and non-periodic ordered orbits. If $\lambda \in \left[\cup_{i=1}^{\infty}[\lambda_i^-, \lambda_i^+]\right]^c$, and $\mathcal{O}(x)$ is the corresponding non-periodic ordered orbit, then $A_\lambda = \overline{\mathcal{O}(x)}$.

(d) Suppose $\lambda, \lambda' \in K$ do not both belong to the same interval $[\lambda_i^-, \lambda_i^+]$, and suppose $\lambda < \lambda'$. Let $\underline{x}$, $\underline{x}'$ be the symbolic codes of the corresponding ordered orbits guaranteed by (b) and (c). Then $\underline{x} < \underline{x}'$ in the lexicographic ordering. $\square$

In Bullett & Sentenac [10], the above proposition is stated in terms of rotation numbers. Each ordered orbit can be given a rotation number $\rho(A)$ (see the remarks prior to Lemma 3·15), and Proposition 3·20 implies that the map $\lambda \mapsto \rho(A_\lambda)$ is a devil's staircase. A devil's staircase is a continuous, weakly monotonic map, which is locally constant on a set of full measure but not globally constant. On the intervals $[\lambda_i^-, \lambda_i^+]$ of local constancy, the rotation number is *rational* (corresponding to a *periodic* ordered orbit). This phenomenon of 'mode-locking' of rotation numbers at rational values is well-known in the theory of degree one circle maps (see page 392 of Katok & Hasselblatt [25] for the result for homeomorphisms, or Newhouse et al. [43] for the generalisation to arbitrary degree one maps). The connection to our situation arises because each A_λ lies in a semi-circle, and can therefore be considered as an orbit of a certain degree one map (see Boyland [9] or Veerman [70] for details).

Section 3·8. Analysis of $\partial\Omega$.

We want to study the topological boundary $\partial\Omega$ of the convex set Ω. We will consider separately the two symmetric halves of $\partial\Omega$, thinking of them as the graphs of two real-valued functions. Let $\partial\Omega = \partial\Omega^+ \cup \partial\Omega^-$, where

$$\partial\Omega^+ = \{w \in \partial\Omega : \mathrm{Im}(z) \geq 0\} \qquad \text{and} \qquad \partial\Omega^- = \{w \in \partial\Omega : \mathrm{Im}(z) \leq 0\} .$$

Let $g^+ : \Omega \cap \mathbf{R} \to \mathbf{R}$ be the function whose graph is $\partial\Omega^+$ (thinking of $\partial\Omega^+$ as lying in $\mathbf{R}^2$). The convexity of Ω means that g^+ is a concave function. Since g^+ is defined on the closed interval $\Omega \cap \mathbf{R}$, this implies that g^+ is Lipschitz. In fact, the concavity of g^+ allows us to assert the following (see Royden [60], page 113). The right and left derivatives of g^+ exist at every interior point of $\Omega \cap \mathbf{R}$, and are both monotone decreasing functions. At each point, the left-hand derivative is greater than or equal to the right-hand derivative, and in fact they are equal to each other except on a countable set. Analogous differentiability properties hold for the convex function $g^- = -g^+$ whose graph is $\partial\Omega^-$. In particular, g^+ and g^- both have at most countably many points of non-differentiability.

We say that $w = c + ig^+(c)$ (resp. $w = c + ig^-(c)$) is a **point of differentiability** of $\partial\Omega$ if $c \in \Omega \cap \mathbf{R}$ is a point of differentiability of g^+ (resp. g^-). Otherwise we say that w is a **point of non-differentiability** of $\partial\Omega$. We will also refer to the left and right **gradients** of $\partial\Omega$ at the point w, by which we mean the corresponding derivatives of g^+ (or g^-) at the point c.

This analysis omits a discussion of the possible differentiability of the two points in $\partial\Omega \cap \mathbf{R}$ (we know $1 \in \partial\Omega \cap \mathbf{R}$, and we believe that $-\frac{1}{2} \in \partial\Omega \cap \mathbf{R}$). We define $w \in \partial\Omega \cap \mathbf{R}$ to be differentiable if g^+ and g^- both have infinite (one-sided) derivatives at w. From the graphical plot in Appendix D we believe that these two points are *not* differentiable.

From the above discussion, we know that $\partial\Omega$ has at most a countable infinity of points of non-differentiability. Our computation of Ω_{19} (see §3·4, Appendix C, Appendix D), and the fact that many triples of points on $\partial\Omega_{19}$ are close to being collinear, suggest that indeed $\partial\Omega$ is not differentiable, and that if $\mathcal{O}(x)$ is an ordered periodic orbit (there is a countable infinity of such orbits, by the Farey tower construction), then $w = \int z \, d\mu_x(z)$ is a point of non-differentiability of $\partial\Omega$. In §3·11 we formulate this as part (c) of Conjecture III, and Corollary 3·33.

To understand $\partial\Omega$ better we will parametrise it by the circle-valued parameter $\theta \in K$. We can think of θ as indexing a unit normal to $\partial\Omega$ in the $2\pi\theta$ direction. We let $w(\theta)$ denote the point in Ω whose component in the $2\pi\theta$ direction is maximal. That is, $w(\theta) \in \Omega$ satisfies

$$< w(\theta), e^{2\pi i\theta} > \quad \geq \quad < z, e^{2\pi i\theta} > \qquad \text{for all } z \in \Omega. \tag{3·7}$$

Clearly any such $w(\theta)$ must lie in $\partial\Omega$. Moreover, any point of $\partial\Omega$ must be of the form $w(\theta)$, for some $\theta \in K$. We record this as a lemma.

Lemma 3·21. *Suppose $w \in \Omega$. Then $w \in \partial\Omega$ if and only if there exists $\theta \in K$ such that $w = w(\theta)$.* □

Note that (3·7) means we can draw a line of gradient $\tan(2\pi\theta + \pi/2)$ through the point $w(\theta)$, such that all of Ω is on one side of the line. Such a line is a tangent to $\partial\Omega$ at the point w, but note that if w is a point of non-differentiability then the tangent line is not unique. Note that (3·7) does *not* mean that $w(\theta)$ is the point on $\partial\Omega$ which intersects the half-line $\{re^{2\pi i\theta} : r \geq 0\}$.

It is not clear from (3·7) that $w(\theta)$ is uniquely defined. If $\partial\Omega$ contained a line segment normal to the $2\pi\theta$ direction, then every point on the line segment would satisfy (3·7). Thus $w(\theta)$ would be a line segment rather than a single point.

However, we believe (see Corollary 3·30, which relies on Conjecture III in §3·11) that $\partial\Omega$ does *not* contain any line segments. If this is true, then $\theta \mapsto w(\theta)$ is a well-defined parametrisation of $\partial\Omega$. Clearly $\theta \mapsto w(\theta)$ is a continuous parametrisation. In fact, equation (3·9) in §3·9 shows that $w(\theta)$ is related to the Lipschitz functional Q (see §3·3), and it follows that $\theta \mapsto w(\theta)$ is a Lipschitz parametrisation. We now discuss how the regularity of $\theta \mapsto w(\theta)$ relates to the regularity of $\partial\Omega$ itself. We shall see that the places where $\partial\Omega$ is badly behaved correspond to the places where $\theta \mapsto w(\theta)$ is well behaved.

Suppose $w \in \partial\Omega$ is a point of non-differentiability of $\partial\Omega$. This just means that the left and right gradients to $\partial\Omega$ at w do not agree. Therefore $\partial\Omega$ does not have a unique tangent line at the point w. Rather, there is a non-trivial closed interval $[\theta_1, \theta_2]$ such that, for any $\theta \in [\theta_1, \theta_2]$, the line through w of gradient $\tan(2\pi\theta + \pi/2)$ is a tangent to $\partial\Omega$. In other words, $\theta \mapsto w(\theta)$ is constant on the interval $[\theta_1, \theta_2]$. So bad behaviour (non-differentiability) of $\partial\Omega$ corresponds to good behaviour (local constancy) of $\theta \mapsto w(\theta)$.

In §3·11 we conjecture that $\partial\Omega$ has a countable infinity of points of non-differentiability, and that these points are in one-to-one correspondence with the periodic ordered orbits of T. As above, we can associate to each non-differentiable point a closed interval $[\theta_1, \theta_2]$ in parameter space, on which $\theta \mapsto w(\theta)$ is constant. We conjecture (see Conjecture II in §3·10) that the union of these intervals is a set of full Lebesgue measure. This conjecture is analogous to the proved result Proposition 3·20.

Section 3·9. A One-Parameter Family of Trigonometric Functions.

Another approach to studying $\partial\Omega$ and its parametrisation is to use the machinery of §3·3. We have defined $w(\theta)$ to be the point in Ω whose component in the $2\pi\theta$ direction is maximal. If $q(\theta)$ denotes the *size* of that component, then we have

$$q(\theta) = < w(\theta), e^{2\pi i\theta} > . \tag{3·8}$$

Note that $q(\theta)$ is well-defined, even if $w(\theta)$ is a line segment rather than a single point (see the discussion in §3·8).

Let us introduce the family of trigonometric functions $f_\theta : K \to [-1, 1]$ defined by

$$\begin{aligned} f_\theta(x) &= \cos(2\pi x - 2\pi\theta) \\ &= \cos 2\pi x \cos 2\pi\theta + \sin 2\pi x \sin 2\pi\theta \\ &=< e^{2\pi i x}, e^{2\pi i\theta} > . \end{aligned}$$

So $f_\theta(x)$ is just the component of the complex number $e^{2\pi i x}$ in the $2\pi\theta$ direction. If $\mu \in \mathcal{M}$, then $\int f_\theta \, d\mu$ is the component in the $2\pi\theta$ direction of the barycentre $\int_{S^1} z \, d\mu(z)$. Thus $Q(f_\theta) = \sup_{\mu\in\mathcal{M}} \int f_\theta \, d\mu$ is the *largest* component in the $2\pi\theta$ direction of all barycentres. That is,

$$q(\theta) = Q(f_\theta). \tag{3·9}$$

The upshot of this discussion is the following lemma. Recall that maximal measures were defined in Definition 3·9.

Lemma 3·22. *$\int_{S^1} z \, d\mu \in \partial\Omega$ if and only if μ is f_θ-maximal for some $\theta \in K$.*

Proof.

$$\begin{aligned}
\mu \text{ is } f_\theta\text{-maximal for some } \theta &\iff \int f_\theta \, d\mu = Q(f_\theta) \quad \text{for some } \theta \\
&\iff \int f_\theta \, d\mu = q(\theta) \quad \text{for some } \theta \\
&\iff \int_{S^1} z \, d\mu(z) = w(\theta) \quad \text{for some } \theta \\
&\iff \int_{S^1} z \, d\mu(z) \in \partial\Omega \quad \text{(by Lemma 3·21).} \quad \square
\end{aligned}$$

So any information about f_θ-maximal measures will translate into information about measures in $\mathcal{M}(\partial\Omega)$.

Note that each f_θ is clearly Hölder, and not an essential coboundary, so by Proposition 3·14 we know that any f_θ-maximal measure cannot be fully supported. Consequently we deduce

Proposition 3·23. *Any $\mu \in \mathcal{M}(\partial\Omega)$ is not fully supported.*

Proof. Immediate from Proposition 3·14 and Lemma 3·22. $\square$

In §3·10 and §3·11 we make various conjectures about the nature of those measures whose barycentres lie on the boundary $\partial\Omega$. For the remainder of this section we concentrate on those points in the interior of Ω, and show that to each of them we can associate an equilibrium state of a particular kind.

Proposition 3·24. *Suppose m_{tf_θ}, $m_{t'f_{\theta'}}$ are equilibrium states, where $t, t' > 0$, and $\theta, \theta' \in K$, and either $t \neq t'$ or $\theta \neq \theta'$ (or both). Then*

$$\int_{S^1} z \, dm_{tf_\theta}(z) \neq \int_{S^1} z \, dm_{t'f_{\theta'}}(z).$$

Proof. The functions tf_θ, $t'f_{\theta'}$ are not essentially cohomologous, so their corresponding equilibrium states m_{tf_θ}, $m_{t'f_{\theta'}}$ are not equal (see page 226 of Walters [72]).

Now suppose, for a contradiction, that

$$\int_{S^1} z \; dm_{tf_\theta}(z) = w = \int_{S^1} z \; dm_{t'f_{\theta'}}(z). \tag{3·10}$$

Consider the components of w in the $2\pi\theta$ and $2\pi\theta'$ directions. Equation (3·10) gives

$$\int f_\theta \; dm_{tf_\theta} = \int f_\theta \; dm_{t'f_{\theta'}} = < w, e^{2\pi i\theta} > =: b \;, \tag{3·11}$$

and

$$\int f_{\theta'} \; dm_{tf_\theta} = \int f_{\theta'} \; dm_{t'f_{\theta'}} = < w, e^{2\pi i\theta'} > =: b' \;. \tag{3·12}$$

Equation (3·11) means that m_{tf_θ} and $m_{t'f_{\theta'}}$ both belong to $\mathcal{M}(f_\theta, b)$. But Proposition 3·6 (b) implies that $m = m_{tf_\theta}$ is the unique measure in $\mathcal{M}(f_\theta, b)$ satisfying $h(m) = \sup_{\mu\in\mathcal{M}(f_\theta,b)} h(\mu)$. In particular we obtain

$$h(m_{tf_\theta}) > h(m_{t'f_{\theta'}}). \tag{3·13}$$

Similarly, equation (3·12) means that m_{tf_θ} and $m_{t'f_{\theta'}}$ both belong to $\mathcal{M}(f_{\theta'}, b')$. But Proposition 3·6 (b) implies that $m = m_{t'f_{\theta'}}$ is the unique measure in $\mathcal{M}(f_{\theta'}, b')$ satisfying $h(m) = \sup_{\mu\in\mathcal{M}(f_{\theta'},b')} h(\mu)$. In particular we obtain

$$h(m_{t'f_{\theta'}}) > h(m_{tf_\theta}). \tag{3·14}$$

Equations (3·13) and (3·14) together give the required contradiction. □

Lemma 3·25. *Let $\theta \in K$, and let m_{tf_θ} denote the equilibrium state of tf_θ, for any $t \in \mathbf{R}$. Then the map $(t,\theta) \mapsto \int_{S^1} z \; dm_{tf_\theta}(z)$ is continuous.*

Proof. Write

$$\int_{S^1} z \; dm_{tf_\theta}(z) = \int \cos(2\pi x) dm_{tf_\theta}(x) + i \int \sin(2\pi x) dm_{tf_\theta}(x). \tag{3·15}$$

The function tf_θ has an analytic (and hence a Hölder) dependence on (t,θ). Thus m_{tf_θ} depends continuously (with respect to the weak* topology) on (t,θ), by Lemma 3·4.

The definition of weak* continuity means that both real and imaginary parts of (3·15) depend continuously on (t, θ). $\square$

Proposition 3·26. *If $w \in int(\Omega)$ then there exists an equilibrium state m_{tf_θ}, where $t \geq 0$ and $\theta \in K$, such that*

$$w = \int_{S^1} z \; dm_{tf_\theta}(z) \; .$$

Proof. Fix $\theta \in K$ for the moment. Define $\gamma_\theta : [0, \infty) \to \Omega$ by

$$\gamma_\theta(t) = \int_{S^1} z \; dm_{tf_\theta}(z).$$

We know that γ_θ is a continuous curve, by Lemma 3·25. Note that

$$\gamma_\theta(0) = \int_{S^1} z \; dm_0(z) = \int_{S^1} z \; dl(z) = 0 \; ,$$

since Lebesgue measure l is the measure of maximal entropy of T.

By Proposition 3·7 and equation (3·9), we know that

$$\int f_\theta \; dm_{tf_\theta} \to Q(f_\theta) = q(\theta) \qquad \text{as } t \to \infty.$$

This means that the component of the barycentre $\int_{S^1} z \; dm_{tf_\theta}(z)$ in the $2\pi\theta$ direction is converging to the largest possible component in the $2\pi\theta$ direction. Thus the barycentre $\int_{S^1} z \; dm_{tf_\theta}(z)$ is itself converging to the barycentre $w(\theta)$ whose component in the $2\pi\theta$ direction is the largest possible. That is,

$$\gamma_\theta(t) \to w(\theta) \qquad \text{as } t \to \infty. \tag{3·16}$$

Therefore $\Gamma_\theta = \left\{ \int_{S^1} z \; dm_{tf_\theta}(z) : t \geq 0 \right\}$ is a half-open curve in Ω, with closed endpoint at 0 and open endpoint at $w(\theta)$. Recall from the discussion in §3·8 that we have not proved that $w(\theta)$ is a single point. It might be a closed interval instead. In this case (3·16) just means that $\gamma_\theta(t)$ accumulates at some subset of $w(\theta)$. In the following argument, the important point is just that $\theta \mapsto w(\theta)$ is not globally constant. This is clearly the case, as Ω is a two dimensional convex shape.

If we now let $\theta \in K$ vary, then the family γ_θ is continuous in θ, by Lemma 3·25. Since Ω is convex then it is contractible, so that the family of curves Γ_θ 'sweeps out' the whole of $int(\Omega)$. That is,

$$\bigcup_{\theta \in K} \Gamma_\theta = int(\Omega). \quad \square$$

Remark. Proposition 3·24 implies that the family of curves Γ_θ in Proposition 3·26 only intersect each other at $0 \in \mathbf{C}$. It would be interesting to study further the geometry of this family of curves.

Corollary 3·27. *If $w \in int(\Omega)$ then $\mathcal{M}(w)$ contains a unique equilibrium state of the form m_{tf_θ}. This equilibrium state is the unique measure of maximal entropy amongst all measures in $\mathcal{M}(w)$. This maximal entropy is equal to $P(t\ (f_\theta - < w, e^{2\pi i\theta} >))$.*

Proof. Existence of such an equilibrium state follows from Proposition 3·26. Uniqueness follows from Proposition 3·24, and from the observation that letting t take negative values does not give any new equilibrium states (since $tf_\theta = -tf_{\theta+\frac{1}{2}}$).

Suppose $\mu, m_{tf_\theta} \in \mathcal{M}(w)$, where $\mu \neq m_{tf_\theta}$. This implies that

$$\mu, m_{tf_\theta} \in \mathcal{M}(f_\theta, < w, e^{2\pi i\theta} >).$$

But then Proposition 3·6 (b) implies that $h(m_{tf_\theta}) > h(\mu)$.

By Proposition 3·6 (c) we have $h(m_{tf_\theta}) = P(t\ (f_\theta - < w, e^{2\pi i\theta} >))$. $\square$

Corollary 3·28. *If $w \in int(\Omega)$ then $\mathcal{M}(w)$ contains an ergodic measure with positive entropy.*

Proof. This follows from Proposition 3·26, and from the fact that all equilibrium states are ergodic and have positive entropy. $\square$

There are some further remarks on entropy at the end of §3·11.

Section 3·10. Conjectures I and II.

Throughout this section, if $\mathcal{O}(x)$ is an ordered orbit, then $\mu_{\mathcal{O}(x)}$ will denote the unique T-invariant probability measure supported on $\overline{\mathcal{O}(x)}$ (see Proposition 3·17).

Conjecture I. *There is a sequence of disjoint non-trivial intervals $[\theta_i^-, \theta_i^+] \subset K$ such that*

(a) $\cup_{i=1}^{\infty}(\theta_i^-, \theta_i^+)$ has full Lebesgue measure. Its complement in K is a Cantor set of zero Hausdorff dimension.

(b) There is a one-to-one correspondence between intervals in the sequence and periodic ordered orbits. If $[\theta_i^-, \theta_i^+]$ is an interval, then the corresponding periodic ordered orbit $\mathcal{O}(x)$ is strictly maximal for all functions f_θ, $\theta \in [\theta_i^-, \theta_i^+]$, and uniformly strictly maximal for all functions f_θ, $\theta \in (\theta_i^-, \theta_i^+)$.

(c) There is a one-to-one correspondence between points of $\left[\cup_{i=1}^{\infty}[\theta_i^-, \theta_i^+]\right]^c$ and non-periodic ordered orbits. If $\theta \in \left[\cup_{i=1}^{\infty}[\theta_i^-, \theta_i^+]\right]^c$, then the corresponding non-periodic ordered orbit $\mathcal{O}(x)$ is strictly maximal for f_θ.

(d) Suppose $\theta, \theta' \in K$ do not both belong to the same interval $[\theta_i^-, \theta_i^+]$, and suppose $\theta < \theta'$. Let $\underline{x}$, $\underline{x}'$ be the symbolic codes of the corresponding ordered orbits guaranteed by (b) and (c). Then $\underline{x} < \underline{x}'$ in the lexicographic ordering.

The open parameter intervals in part (b) are entirely natural, as the parametrisation $\theta \mapsto f_\theta$ is Hölder (in fact analytic), and the property of uniform strict maximality persists under sufficiently small Hölder perturbations. The substance of Conjecture I is that almost all (but not all) f_θ have a uniformly strictly maximal periodic orbit, and that these orbits are ordered orbits.

Conjecture II. *There is a sequence of disjoint non-trivial intervals $[\theta_i^-, \theta_i^+] \subset K$ such that*

(a) $\cup_{i=1}^{\infty}(\theta_i^-, \theta_i^+)$ has full Lebesgue measure. Its complement in K is a Cantor set of zero Hausdorff dimension.

(b) There is a one-to-one correspondence between intervals in the sequence and periodic ordered orbits. Suppose $[\theta_i^-, \theta_i^+]$ is an interval, and the corresponding periodic ordered orbit is $\mathcal{O}(x)$. Then the map $\theta \mapsto w(\theta)$ is constant on $[\theta_i^-, \theta_i^+]$, with value $\int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z)$, where $\mu_{\mathcal{O}(x)}$ is the T-invariant Borel probability measure concentrated on $\mathcal{O}(x)$. Moreover, $\mathcal{M}(w(\theta)) = \{\mu_{\mathcal{O}(x)}\}$ for all $\theta \in [\theta_i^-, \theta_i^+]$.

(c) There is a one-to-one correspondence between points of $\left[\cup_{i=1}^{\infty}[\theta_i^-, \theta_i^+]\right]^c$ and non-periodic ordered orbits. If $\theta \in \left[\cup_{i=1}^{\infty}[\theta_i^-, \theta_i^+]\right]^c$, and $\mathcal{O}(x)$ is the corresponding non-periodic orbit, then $w(\theta) = \int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z)$, where $\mu_{\mathcal{O}(x)}$ is the T-invariant Borel probability measure concentrated on $\overline{\mathcal{O}(x)}$. Moreover, $\mathcal{M}(w(\theta)) = \{\mu_{\mathcal{O}(x)}\}$.

(d) Suppose $\theta, \theta' \in K$ do not both belong to the same interval $[\theta_i^-, \theta_i^+]$, and suppose $\theta < \theta'$. Let $\underline{x}$, $\underline{x}'$ be the symbolic codes of the corresponding ordered orbits guaranteed by (b) and (c). Then $\underline{x} < \underline{x}'$ in the lexicographic ordering.

We believe that Conjecture I and Conjecture II are equivalent, with the same sequence of intervals in each case, though we cannot prove all parts of this equivalence. We outline below which parts we can and cannot prove.

I(a) $\Longleftrightarrow$ II(a) and **I(d) $\Longleftrightarrow$ II(d)** are clear.

I(b) $\Rightarrow$ II(b). Since the periodic ordered orbit $\mathcal{O}(x)$ is strictly maximal for f_θ, $\theta \in [\theta_i^-, \theta_i^+]$, then by Lemma 3·9 (a) we deduce that $\mu_{\mathcal{O}(x)}$ is f_θ-maximal for $\theta \in [\theta_i^-, \theta_i^+]$. This implies (see the proof of Lemma 3·22, for example) that $w(\theta) = \int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z)$ for $\theta \in [\theta_i^-, \theta_i^+]$. Since $\mathcal{O}(x)$ is uniformly strictly maximal for f_θ, $\theta \in (\theta_i^-, \theta_i^+)$, then Proposition 3·13 tells us that $\mu_{\mathcal{O}(x)}$ is the *unique* f_θ-maximal measure for $\theta \in (\theta_i^-, \theta_i^+)$. We believe that $\mu_{\mathcal{O}(x)}$ is also the unique f_θ-maximal measure for the endpoints $\theta = \theta_i^-$ and $\theta = \theta_i^+$, but we cannot prove this.

II(b) $\Rightarrow$ I(b). Since $\mathcal{M}(w(\theta)) = \{\mu_{\mathcal{O}(x)}\}$ for $\theta \in [\theta_i^-, \theta_i^+]$, then $\mu_{\mathcal{O}(x)}$ is the unique f_θ-maximal measure for $\theta \in [\theta_i^-, \theta_i^+]$. By Lemma 3·9 (b) we deduce that $\mathcal{O}(x)$ is strictly maximal for f_θ, $\theta \in [\theta_i^-, \theta_i^+]$. However, we cannot prove that $\mathcal{O}(x)$ is *uniformly* strictly maximal for f_θ, $\theta \in (\theta_i^-, \theta_i^+)$.

I(c) $\Rightarrow$ II(c). Since $\mathcal{O}(x)$ is strictly maximal for f_θ, then $\mu_{\mathcal{O}(x)}$ is certainly an f_θ-

maximal measure, by Lemma 3·9 (a). Therefore $w(\theta) = \int_{S^1} z \; d\mu_{\mathcal{O}(x)}(z)$. However, we cannot prove that $\mu_{\mathcal{O}(x)}$ is the *unique* measure in $\mathcal{M}(w(\theta))$.

II(c) $\Rightarrow$ I(c). Since $\mathcal{M}(w(\theta)) = \{\mu_{\mathcal{O}(x)}\}$, then $\mu_{\mathcal{O}(x)}$ is the unique f_θ-maximal measure. By Lemma 3·9 (b) we deduce that $\mathcal{O}(x)$ is strictly maximal for f_θ.

Conjecture II is reminiscent of the proved result Proposition 3·20, since for each parameter value θ we pick out (the closure of) some ordered orbit. In both results every *periodic* ordered orbit corresponds to a non-trivial interval of constancy in parameter space (this phenomenon is similar to the 'mode-locking' of circle maps mentioned in §3·7). The union of such intervals has full Lebesgue measure, and its complement has zero Hausdorff dimension. The fact that Proposition 3·20 is true suggests that Corollary II is true as well.

It is natural to ask whether the intervals in Conjecture II are the *same* as those in Proposition 3·20. This is clearly not the case. For example, the fixed point $0 \in K$ is the unique ordered orbit contained in the semi-circle $[\lambda - \frac{1}{4}, \lambda + \frac{1}{4}]$, for all $\lambda \in [\frac{3}{4}, \frac{1}{4}]$. Thus the map $\lambda \mapsto A_\lambda$ is constant on the interval $[\frac{3}{4}, \frac{1}{4}]$. In contrast, the map $\theta \mapsto w(\theta)$ is clearly *not* constant on the whole of the interval $[\frac{3}{4}, \frac{1}{4}]$. To see this just note that the fixed point $x = 0$ maximises the function $f_0(x) = \cos 2\pi x$, while it certainly does not maximise $f_{\frac{1}{4}}(x) = \sin 2\pi x$. Thus $w(0) = e^{2\pi i 0} = 1$, but $w(\frac{1}{4}) \neq 1$.

It was calculated that the (symmetric) interval of constancy around $\theta = 0$ is approximately $[-0.14955, 0.14955]$.

In general the intervals of constancy of $\theta \mapsto w(\theta)$ are not the same as those of $\lambda \mapsto A_\lambda$.

Bullett & Sentenac [10] show that if an ordered periodic orbit has (rational) rotation number p/q (in lowest terms), then the corresponding interval of constancy has length $\frac{1}{2(2^q-1)}$. We know of no such formula for the lengths of the intervals of constancy of the map $\theta \mapsto w(\theta)$.

However, we can *estimate* the intervals of constancy corresponding to the first five ordered orbits generated by the Farey construction of §3·6. These estimates are accurate to 6 decimal places, and were obtained using the Mathematica computer program. The intermediate calculations were performed with an accuracy of 100 decimal places. The technique used was the following.

1. Choose the ordered periodic orbit $\mathcal{O}(x)$, and let $\underline{x}$ be its symbolic code.

2. Compute the barycentre w corresponding to the measure supported on $\mathcal{O}(x)$.

3. Let $\underline{x}^-$ (resp. $\underline{x}^+$) be an adjacent code to $\underline{x}$ on some j^{th} Farey level (where typically we took $j \geq 12$) such that $\underline{x}^- < \underline{x}$ (resp. $\underline{x} < \underline{x}^+$).

4. Compute the barycentre w^- (resp. w^+) corresponding to the measure supported on $\mathcal{O}(x^-)$ (resp. $\mathcal{O}(x^+)$).

5. Compute the gradient d^- (resp. d^+) of the straight line through the points w and w^- (resp. w^+). These represent the (approximate) derivatives at the point $w \in \partial\Omega$.

6. Let

$$\theta^- = \frac{\tan^{-1}(d^-) + \pi/2}{2\pi}$$

and

$$\theta^+ = \frac{\tan^{-1}(d^+) + \pi/2}{2\pi} .$$

7. The interval $[\theta^-, \theta^+]$ is the approximate interval of constancy corresponding to $\mathcal{O}(x)$.

The approximate intervals of constancy of $\theta \mapsto w(\theta)$, corresponding to the first five codes in the Farey construction (see §3·6), are as follows. We also list their lengths. Note that the intervals corresponding to the conjugate codes (i.e. obtained by swapping 0's and 1's) are obtained by reflecting in the mid-point $\frac{1}{2}$ of our parameter circle K. The codes 0 and 01, whose corresponding orbits are symmetric in the circle, have intervals of constancy which are also symmetric.

Code	Interval of Constancy	Length
0	$[-0.149550,\ 0.149550]$	0.2991
01	$[0.420148,\ -0.420148]$	0.159704
001	$[0.279199,\ 0.367215]$	0.088016
0001	$[0.216946,\ 0.266213]$	0.049267
00101	$[0.374417,\ 0.404815]$	0.030398

Section 3·11. Conjecture III.

Conjecture III.

(a) There is a one-to-one correspondence between the points of $\partial\Omega$ and the ordered orbits of T. The correspondence is given by $\mathcal{O}(x) \mapsto \int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z)$, where $\mu_{\mathcal{O}(x)}$ is the unique T-invariant Borel probability measure concentrated on the closure of the ordered orbit $\mathcal{O}(x)$.

(b) If $w \in \partial\Omega$ then $\mathcal{M}(w)$ is a singleton set.

(c) $w \in \partial\Omega$ is a point of non-differentiability of $\partial\Omega$ if and only if its corresponding ordered orbit is periodic.

(d) Suppose $w, w' \in \partial\Omega$ satisfy $Arg(w) < Arg(w')$, where the principal argument Arg lies in the range $[0, 2\pi)$. If $\underline{x}$, $\underline{x}'$ are the symbolic codes of the corresponding ordered orbits guaranteed by (a), then $\underline{x} < \underline{x}'$ in the lexicographic ordering.

It is a simple check that Conjecture III is equivalent to parts (b), (c) and (d) of Conjecture II. Part (a) of Conjecture II implies that as the tangent line to $\partial\Omega$ moves around anticlockwise, it 'jumps' (rather than varies continuously) through a full angular measure of 2π.

We now present some corollaries of Conjecture III (and therefore of Conjecture II).

Corollary 3·29. *(Assuming Conjecture III)*

Every point on $\partial\Omega$ is an extremal point of Ω.

Proof. Suppose $w \in \partial\Omega$. By part (b) of Conjecture III we know that $\mathcal{M}(w)$ consists of a single measure μ. By part (a) of Conjecture III we know that μ is concentrated on the closure of some ordered orbit, so Corollary 3·18 implies that μ is ergodic. Then by Lemma 3·2 we deduce that $w \in E(\Omega)$. $\square$

Corollary 3·30. *(Assuming Conjecture III)*

There are no line segments in $\partial\Omega$.

Proof. This is immediate from Corollary 3·29. □

Corollary 3·31. *(Assuming Conjecture III)*

The parametrisation $\theta \mapsto w(\theta)$ of $\partial\Omega$ is well-defined.

Proof. In the discussion after Lemma 3·21 we remarked that this parametrisation is well-defined provided there are no line segments in $\partial\Omega$. The result follows from Corollary 3·30. □

Corollary 3·32. *(Assuming Conjecture III)*

The countable set

$$\left\{ \int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z) : \mathcal{O}(x) \text{ is a periodic ordered orbit} \right\}$$

is a dense subset of $\partial\Omega$.

The uncountable set

$$\left\{ \int_{S^1} z \, d\mu_{\mathcal{O}(x)}(z) : \mathcal{O}(x) \text{ is a non-periodic ordered orbit} \right\}$$

is a dense subset of $\partial\Omega$.

Proof. Suppose $w \in \partial\Omega$. Pick $w' \in \partial\Omega$ arbitrarily close to w, with $\mathrm{Arg}(w) < \mathrm{Arg}(w')$. Suppose $\underline{x}$, $\underline{x}'$ are the symbolic codes of the corresponding ordered orbits. Then part (d) of Corollary B gives us $\underline{x} < \underline{x}'$.

From the Farey tower construction (see §3·6) it is clear that between $\underline{x}$ and $\underline{x}'$ (in the lexicographic ordering) there are infinitely many finite Sturmian codes (corresponding to periodic ordered orbits), and infinitely many infinite Sturmian codes (corresponding to non-periodic ordered orbits). So choose a finite Sturmian code $\underline{y^1}$ with $\underline{x} < \underline{y^1} < \underline{x}'$, and an infinite Sturmian code $\underline{y^2}$ with $\underline{x} < \underline{y^2} < \underline{x}'$. By part (d) of Conjecture III we have

$$\mathrm{Arg}(w) < \mathrm{Arg}(y^1) < \mathrm{Arg}(w')$$

and

$$\mathrm{Arg}(w) < \mathrm{Arg}(y^2) < \mathrm{Arg}(w'). \quad \square$$

Corollary 3·33. *(Assuming Conjecture III)*

The boundary $\partial\Omega$ has a countable infinity of points of non-differentiability. These points are dense in $\partial\Omega$.

Proof. By part (c) of Conjecture III we know that the points of non-differentiability of $\partial\Omega$ correspond to the periodic ordered orbits. There is a countable infinity of periodic ordered orbits, and by Corollary 3·32 their barycentres are densely embedded in $\partial\Omega$. $\square$

Corollary 3·34. *(Assuming Conjecture III)*

All measures $\mu \in \mathcal{M}(\partial\Omega)$ are ergodic and have zero entropy.

Proof. By parts (a) and (b) of Conjecture III we know that $\mathcal{M}(\partial\Omega)$ is the union of those T-invariant Borel probability measures which are concentrated on the closure of some ordered orbit. By Proposition 3·18 we know that all such measures are ergodic and have zero entropy. $\square$

Corollary 3·35. *(Assuming Conjecture III)*

For all $w \in \Omega$ there exists an ergodic measure $\mu \in \mathcal{M}$ such that $w = \int_{S^1} z \, d\mu(z)$.

Proof. By Proposition 3·28 we know that if w is an interior point of Ω, then $\mathcal{M}(w)$ contains an ergodic measure. By Corollary 3·34 we know that if w is on the boundary of Ω then $\mathcal{M}(w)$ is a singleton set containing an ergodic measure. $\square$

Remark. We can define an entropy function $H : \Omega \to [0, \log 2]$ by

$$H(w) = \sup_{\mu \in \mathcal{M}(w)} h(\mu).$$

From Corollary 3·27 we deduce that H is continuous on $int(\Omega)$. The point $0 \in \Omega$ is the unique global maximum of H, with $H(0) = \log 2$. This is because Lebesgue measure has barycentre 0, and is the unique measure of maximal entropy of T. By Corollary 3·28 we know that $H > 0$ on the interior of Ω. If Conjecture III is true then by Corollary 3·34 we also know that H is identically zero on $\partial\Omega$. Therefore $H^{-1}(0) = \partial\Omega$. It would be interesting to study the structure of the other equipotentials $H^{-1}(c)$.

Section 3·12. The Positive Analytic Livsic Conjecture.

The following result was proved, in a more general context, by Sharp [65].

Proposition 3·36. *(Sharp, [65]) Let $F : K \to \mathbf{R}$ be a continuous function such that $\int F d\mu \geq 0$ for all $\mu \in \mathcal{M}$. Then there exists a continuous function $F' : K \to \mathbf{R}$ such that*

(a) $F' \geq 0$

(b) $\int F' d\mu = \int F d\mu$ for all $\mu \in \mathcal{M}$. □

Mark Pollicott and Richard Sharp wondered (personal communication) if there was an analytic version of the above result. We call this the Positive Analytic Livsic Conjecture. The precise statement is the following.

Conjecture IV. Let $U \subset \mathbf{C}$ be an open annulus containing S^1. Let $F : U \to \mathbf{R}$ be an analytic function such that $\int_{S^1} F d\mu \geq 0$ for all $\mu \in \mathcal{M}$. Then there exists an analytic function $F' : U \to \mathbf{R}$ such that

(a) $F' \geq 0$

(b) $\int_{S^1} F' d\mu = \int_{S^1} F d\mu$ for all $\mu \in \mathcal{M}$.

We believe this conjecture to be false, for if it were true then Conjectures I, II and III would be false, by the following reasoning due to Pollicott and Sharp (personal communication).

Proposition 3·37. *Conjecture IV implies that every f_θ-maximal measure has finite support.*

Proof. Let $\theta \in K$ be arbitrary, and suppose $m \in \mathcal{M}$ is an f_θ-maximal measure. So $\int f_\theta dm = Q(f_\theta)$. Define $G_\theta : K \to \mathbf{R}$ by

$$\begin{aligned} G_\theta(x) &= Q(f_\theta) - f_\theta(x) \\ &= Q(f_\theta) - \cos(2\pi x - 2\pi\theta). \end{aligned}$$

Recall that $\psi(x) = e^{2\pi i x}$ maps K homeomorphically onto S^1, and note that $F_\theta = G_\theta \circ \psi^{-1}$ satisfies the hypotheses of Conjecture IV. If this conjecture is true then there exists an analytic function $F'_\theta \geq 0$ such that $\int_{S^1} F'_\theta \, d\mu = \int_{S^1} F_\theta \, d\mu$ for all $\mu \in \mathcal{M}$.

In particular we obtain

$$\int_{S^1} F'_\theta \, dm = \int_{S^1} F_\theta \, dm = 0.$$

Since F'_θ is non-negative, this implies that its set of zeros is given full measure by m. But since F'_θ is analytic and not identically zero, this set of zeros must be finite. Therefore m is supported on a finite set (contained in the zero set of F'_θ). $\square$

Proposition 3·38. *Conjectures I and IV are incompatible*

Proof. Conjecture I says that for certain parameter values θ, we have an f_θ-maximal measure supported on an infinite set (namely the orbit closure of a non-periodic ordered orbit). By Proposition 3·37 we deduce that Conjecture I and Conjecture IV are incompatible. $\square$

Proposition 3·39. *Conjecture IV implies that $\partial\Omega$ contains line segments.*

Proof. Suppose Conjecture IV is true. By Proposition 3·37, Conjecture IV implies that every f_θ-maximal measure has finite support. By Lemma 3·22 this means that every measure in $\mathcal{M}(\partial\Omega)$, and hence every measure in $\mathcal{M}(E(\Omega))$, has finite support. If $w \in E(\Omega)$ then Lemma 3·3 guarantees us an ergodic measure $m \in \mathcal{M}(w)$, which must have finite support, by the above. But there are only countably many finitely supported ergodic measures, since each one is concentrated on a single periodic orbit. Therefore $E(\Omega)$ is a countable set. But $\partial\Omega$ is uncountable, so there exist non-extremal points on $\partial\Omega$. Now any non-extremal point on $\partial\Omega$ lies on a line segment joining two extremal points of $\partial\Omega$, so $\partial\Omega$ just consists of a countable number of line segments. $\square$

Proposition 3·40. *Conjectures III and IV are incompatible.*

Proof. Conjecture III implies that $\partial\Omega$ does not contain any line segments (see Corollary 3·30). By Proposition 3·39 we deduce that Conjectures III and IV are incompatible. $\square$

Proposition 3·41. *Conjectures II and IV are incompatible.*

Proof. Conjectures II and III are equivalent (see §3·11). The result follows from Proposition 3·40. □

Appendix A

In this appendix we present a proof of Lemma 1·22.

Throughout the appendix we will use a central dot '$\cdot$' to denote composition of maps. We will use a low dot '.' for products of matrices, vectors and numbers, whenever this aids clarity. For compactness of notation we will always use a lower case d to denote derivatives (in contrast to the upper case D used elsewhere).

Let $\{T_i\}_{i\in I}$ be the family of inverse branches of T, indexed by the finite set I. Recall that the contraction constant $\gamma < 1$ is such that $\|dT_i\|_\infty \le \gamma < 1$ for each $i \in I$.

Then, provided the composition $T_{i_n} \cdots T_{i_1}$ is well-defined, it will be a local inverse branch of the n-fold composition $T^n = T \cdots T$.

For notational convenience we will introduce the vector $\underline{i} = (i_n, \ldots, i_1)$. The length of this vector will be written $|\underline{i}| = n$. We will write $T_{\underline{i}} = T_{i_n} \cdots T_{i_1}$. It will be important not to confuse the single inverse T_i with the n-fold inverse $T_{\underline{i}}$.

Recall (see Definition 1·6) that the matrix operator $L_\theta : C^k(X, \mathbb{C}^{d^2}) \to C^k(X, \mathbb{C}^{d^2})$ is given by the formula

$$\begin{aligned} L_\theta w(x) &= \sum_{Ty=x} g(y)\theta(y)w(y) \\ &= \sum_{Ty=x} \psi(y)w(y), \end{aligned}$$

where we define $\psi(z) = g(z)\theta(z)$.

If we introduce the notation

$$g_n(z) = g(T^{n-1}z)g(T^{n-2}z)\ldots g(Tz)g(z),$$

$$\theta_n(z) = \theta(T^{n-1}z)\theta(T^{n-2}z)\ldots \theta(Tz)\theta(z),$$

$$\psi_n(z) = \psi(T^{n-1}z)\psi(T^{n-2}z)\ldots \psi(Tz)\psi(z),$$

then we can express the n^{th} iterate of L_θ as

$$\begin{aligned} L_\theta^n w(x) &= \sum_{|\underline{i}|=n} \psi_n(T_{\underline{i}}x)w(T_{\underline{i}}x) \\ &= \sum_{|\underline{i}|=n} g_n(T_{\underline{i}}x)\theta_n(T_{\underline{i}}x)w(T_{\underline{i}}x). \end{aligned} \tag{4·1}$$

Note also that the n^{th} iterate of the normalised Ruelle-Perron-Frobenius operator $\bar{L} : C^{\mathbf{k}}(X, \mathbb{C}) \to C^{\mathbf{k}}(X, \mathbb{C})$ (see Definition 1·4) can be expressed as

$$\bar{L}^n w(x) = \sum_{|\underline{i}|=n} g_n(T_{\underline{i}} x) w(T_{\underline{i}} x). \tag{4·2}$$

We want to prove the following lemma.

Lemma 1·22. *Suppose $w \in C^{\mathbf{k}}(X, \mathbb{C}^{d^2})$, where $0 < \mathbf{k} = (k, \epsilon) \leq \mathbf{r} - 1$. Then for any $\gamma_0 \in (\gamma, 1)$ there exists $C_1 > 0$ such that for all $n \geq 0$:*

$$\| L_\theta^n w \|_k \leq C_1 \sum_{j=0}^{k} \| d^j w \|_\infty \gamma_0^{nj}.$$

Proof. By equation (4·1) we have

$$L_\theta^n w(x) = \sum_{|\underline{i}|=n} \psi_n(T_{\underline{i}} x) w(T_{\underline{i}} x).$$

For any $0 \leq j \leq k$ we can use Leibniz's Theorem (see Field [16], page 163) to differentiate this expression j times, and then estimate by the supremum norm to obtain

$$\| d^j(L_\theta^n w) \|_\infty \leq \sum_{|\underline{i}|=n} \sum_{t=0}^{j} \binom{j}{t} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \| d^t(w \cdot T_{\underline{i}}) \|_\infty . \tag{4·3}$$

To prove the lemma, we would like to estimate the terms

$$\| d^t(w \cdot T_{\underline{i}}) \|_\infty \quad \text{and} \quad \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty .$$

To this end we have the following two lemmas.

Lemma A. *If $1 \leq t \leq k$ then there exists a constant $E_t > 0$ (independent of n) such that for all $n \geq 0$ we have*

$$\| d^t(w \cdot T_{\underline{i}}) \|_\infty \leq E_t \sum_{l=1}^{t} \| d^l w \|_\infty \gamma_0^{nl}.$$

Lemma B. *For any $j-t \in \{0,\dots,j\}$ there exists a constant $k_{j-t} > 0$ (independent of n) such that for all $n \geq 0$ we have*

$$\sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \leq k_{j-t}.$$

The idea behind the proof of both lemmas is the same. We obtain expressions for $d^m T_{\underline{i}}$ and $d^{j-t}(\psi_n \cdot T_{\underline{i}})$ in terms of derivatives of local inverses T_i (and of ψ). We find that in both expressions the *higher order* derivatives of the T_i (and of ψ) are swamped by the abundance of *first order* derivatives of the T_i. We bound the first order derivatives dT_i by the contractive constant γ, while we estimate all other derivatives by the supremum norm.

Proof of Lemma A.

We want to calculate the t^{th} derivative of the composition $w \cdot T_{\underline{i}}$. Using the explicit formula for this (see Field [16], page 164) and estimating by the supremum norm, we obtain

$$\| d^t(w \cdot T_{\underline{i}}) \|_\infty \leq \sum_{l=1}^{t} \| d^l w \|_\infty \Big(\sum_{\substack{l_1+2l_2+\ldots+tl_t=t \\ l_1+l_2+\ldots+l_t=l}} A_{t,l,l_1,\ldots,l_t} \| dT_{\underline{i}} \|_\infty^{l_1} \| d^2 T_{\underline{i}} \|_\infty^{l_2} \cdots \| d^t T_{\underline{i}} \|_\infty^{l_t} \Big)$$

where each $A_{t,l,l_1,\ldots,l_t} \in \mathbb{N}$ is an explicit combinatorial constant.

We now want to estimate each $\| d^m T_{\underline{i}} \|_\infty$, for $m \geq 1$. We do this by applying the chain rule to the expression $T_{\underline{i}} = T_{i_n} \cdots T_{i_1}$, and then continually applying the product rule for derivatives and the chain rule.

We obtain an expression for $d^m T_{\underline{i}}$, and claim that the number of terms in this expression grows polynomially with n. In fact we claim there exist polynomials $P_m(n)$ in n, of degree $m-1$, such that

$$\| d^m T_{\underline{i}} \|_\infty \leq P_m(n)\gamma^n. \tag{4·4}$$

In the case $m=1$ we apply the chain rule to the composition $T_{\underline{i}} = T_{i_n} \cdots T_{i_1}$ to obtain

$$dT_{\underline{i}} = dT_{i_n} \cdot (T_{i_{n-1}} \cdots T_{i_1}).dT_{i_{n-1}} \cdot (T_{i_{n-2}} \cdots T_{i_1}) \ldots dT_{i_2} \cdot T_{i_1}.dT_{i_1}. \tag{4·5}$$

Therefore we can estimate

$$\| dT_{\underline{i}} \|_\infty \leq \| dT_{i_n} \|_\infty \cdots \| dT_{i_1} \|_\infty \leq \gamma^n.$$

So we can take $P_1(n) = 1$.

In the case $m = 2$ we can apply the product rule (followed by the chain rule) to equation (4·5) to obtain

$$\begin{aligned}
&d^2T_{\underline{i}} = \\
&[d^2T_{i_n} \cdot (T_{i_{n-1}} \cdots T_{i_1}).dT_{i_{n-1}} \cdot (T_{i_{n-2}} \cdots T_{i_1}) \ldots dT_{i_1}].dT_{i_{n-1}} \cdot (T_{i_{n-2}} \cdots T_{i_1}) \ldots dT_{i_2} \cdot T_{i_1}.dT_{i_1} \\
&+ \\
&\vdots \\
&+dT_{i_n} \cdot (T_{i_{n-1}} \cdots T_{i_1}).dT_{i_{n-1}} \cdot (T_{i_{n-2}} \cdots T_{i_1}) \ldots dT_{i_2} \cdot T_{i_1}.d^2T_{i_1},
\end{aligned}$$

so that

$$\begin{aligned}
\| d^2T_{\underline{i}} \|_\infty &\leq B_2\gamma^{n-1}(\gamma^{n-1} + \gamma^{n-2} + \ldots + \gamma + 1) \\
&\leq B_2\gamma^{n-1}n,
\end{aligned}$$

where

$$B_2 = \max_{i \in I} \| d^2T_i \|_\infty \, .$$

(Note that in the above we made the very crude approximation that $\gamma^c \leq 1$ for any $c \geq 0$). So we can take $P_2(n) = B_2\gamma^{-1}n$.

In fact for any $m \geq 2$ we can use the following argument. For $|\underline{i}| = n$ we compute $d^mT_{\underline{i}}$ by repeated application of the chain and product rule to equation (4·5). We obtain an expression for $d^mT_{\underline{i}}$ consisting of the sum of $a(n,m)$ terms, where each term is a product of at most $b(n,m)$ factors. The following recurrence relations hold:

$$a(n, m+1) \leq a(n,m)b(n,m) \tag{4·6}$$

$$b(n, m+1) \leq b(n,m) + n - 1. \tag{4·7}$$

(4·6) is true by the product rule for derivatives, since each term splits into as many terms as it has factors. (4·7) is true by the chain rule, since at most $n-1$ new factors are added to any given term, and this occurs if and only if one of the factors of the term is a composition of n maps.

From (4·5) we obtain the initial conditions $a(n,1) = 1$ and $b(n,1) = n$. Thus (4·7) gives

$$\begin{aligned} b(n,m) &\le n + (m-1)(n-1) \\ &\le n + (m-1)n = mn. \end{aligned}$$

Substituting into (4·6) gives

$$a(n,m) \le (m-1)!\; n^{m-1}.$$

Since at most m of the factors in any term (in the expression for $d^m T_{\underline{i}}$) are not of the form dT_i, the supremum norm of any term is bounded above by $B_{(m)}^m \gamma^{b(n,m)-m}$, where

$$B_{(m)} = \max\{\|d^l T_i\|_\infty : 2 \le l \le m,\ i \in I\}.$$

Adding up all terms we can estimate

$$\begin{aligned} \|d^m T_{\underline{i}}\|_\infty &\le a(n,m) B_{(m)}^m \gamma^{b(n,m)-m} \\ &\le (m-1)!\; n^{m-1} B_{(m)}^m \gamma^{mn-m} \\ &\le (m-1)!\; n^{m-1} B_{(m)}^m \gamma^{n-m}. \end{aligned}$$

Taking our degree-$(m-1)$ polynomial $P_m(n) = (m-1)!\; B_{(m)}^m \gamma^{-m} n^{m-1}$, we have proved the claimed formula (4·4).

Since $\gamma_0 \in (\gamma, 1)$, then by (4·4) we can choose constants $C_m > 0$ such that, for all $m \ge 1$,

$$\| d^m T_{\underline{i}} \|_\infty \le C_m \gamma_0^n. \tag{4·8}$$

Substituting the estimates (4·8) into our estimate for $\| d^t(w \cdot T_{\underline{i}}) \|_\infty$ we obtain

$$\begin{aligned} \| d^t(w \cdot T_{\underline{i}}) \|_\infty &\le \sum_{l=1}^{t} \| d^l w \|_\infty \Big(\sum_{\substack{l_1+2l_2+\ldots+tl_t=t \\ l_1+l_2+\ldots+l_t=l}} A_{t,l,l_1,\ldots,l_t} (C_{l_1}\gamma_0^n)^{l_1} \ldots (C_{l_t}\gamma_0^n)^{l_t} \Big) \\ &= \sum_{l=1}^{t} \| d^l w \|_\infty \gamma_0^{nl} \Big(\sum_{\substack{l_1+2l_2+\ldots+tl_t=t \\ l_1+l_2+\ldots+l_t=l}} A_{t,l,l_1,\ldots,l_t} C_{l_1}^{l_1} \ldots C_{l_t}^{l_t} \Big) \\ &\le E_t \sum_{l=1}^{t} \| d^l w \|_\infty \gamma_0^{nl} \end{aligned}$$

where

$$E_t = \max\{ \sum_{\substack{l_1+2l_2+\ldots+tl_t=t \\ l_1+l_2+\ldots+l_t=l}} A_{t,l,l_1,\ldots,l_t} C_{l_1}^{l_1} \ldots C_{l_t}^{l_t} \ : \ 1 \leq l \leq t\}.$$

This completes the proof of Lemma A. $\square$

Lemma B. *There exists a constant $k_{j-t} > 0$ (independent of n) such that for all $n \geq 0$ we have*

$$\sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \leq k_{j-t}.$$

Proof of Lemma B.

If $j - t = 0$ then we have

$$\begin{aligned}
\sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty &= \sum_{|\underline{i}|=n} \| (\psi_n \cdot T_{\underline{i}}) \|_\infty \\
&= \sum_{|\underline{i}|=n} \| (g_n \cdot T_{\underline{i}}).(\theta_n \cdot T_{\underline{i}}) \|_\infty \\
&\leq \sum_{|\underline{i}|=n} \sup_{x\in X} g_n(T_{\underline{i}}(x)). \| \theta_n \cdot T_{\underline{i}} \|_\infty \\
&= \sum_{|\underline{i}|=n} \sup_{x\in X} g_n(T_{\underline{i}}(x)) \quad \text{by Lemma 1·17} \\
&= \sum_{|\underline{i}|=n} g_n(T_{\underline{i}}(z)) \quad \text{for some } z \in X, \text{ since } X \text{ is compact} \\
&= \bar{L}^n 1 \quad \text{by (4·2)} \\
&= 1 \quad \text{by Theorem 1·15 (i)} \\
&=: k_0.
\end{aligned}$$

If $j - t = 1$ then by applying the product and chain rules we obtain

$$\sum_{|\underline{i}|=n} d^{j-t}(\psi_n \cdot T_{\underline{i}}) = \sum_{|\underline{i}|=n} d(\psi_n \cdot T_{\underline{i}})$$

$$= \sum_{|\underline{i}|=n} d[g \cdot T_{i_1} \dots g \cdot (T_{i_n} \cdots T_{i_1}).\theta \cdot T_{i_1} \dots \theta \cdot (T_{i_n} \cdots T_{i_1})]$$

$$= \sum_{|\underline{i}|=n} \Big[(dg) \cdot T_{i_1}.dT_{i_1}.g \cdot (T_{i_2} \cdot T_{i_1}) \dots g \cdot (T_{i_n} \cdots T_{i_1}).\theta \cdot T_{i_1} \dots \theta \cdot (T_{i_n} \cdots T_{i_1})$$

$$+$$

$$\vdots \tag{4·9}$$

$$+ g \cdot T_{i_1} \dots dg \cdot (T_{i_n} \cdots T_{i_1}).dT_{i_n} \cdot (T_{i_{n-1}} \cdots T_{i_1}) \dots dT_{i_1}.\theta \cdot T_{i_1} \dots \theta \cdot (T_{i_n} \cdots T_{i_1})$$

$$+ g \cdot T_{i_1} \dots g \cdot (T_{i_n} \cdots T_{i_1}).d\theta \cdot T_{i_1}.dT_{i_1}.\theta \cdot (T_{i_1} \cdot T_{i_2}) \dots \theta \cdot (T_{i_n} \cdots T_{i_1})$$

$$+$$

$$\vdots$$

$$+ g \cdot T_{i_1} \dots g \cdot (T_{i_n} \cdots T_{i_1}).\theta \cdot T_{i_1} \dots d\theta \cdot (T_{i_n} \cdots T_{i_1}).dT_{i_n} \cdot (T_{i_{n-1}} \cdots T_{i_1}) \dots dT_{i_1}\Big].$$

Since g is a positive *real*-valued function, we can deduce that

$$\sum_{|\underline{i}|=n} \| d(\psi_n \cdot T_{\underline{i}}) \|_\infty$$

$$\leq \sum_{|\underline{i}|=n} \Big[g_n \cdot T_{\underline{i}}(z) \| dg \|_\infty .\|1/g\|_\infty .1^n.\gamma$$

$$+$$

$$\vdots$$

$$+ g_n \cdot T_{\underline{i}}(z) \| dg \|_\infty .\|1/g\|_\infty .1^n.\gamma^n \tag{4·10}$$

$$+ g_n \cdot T_{\underline{i}}(z) \| d\theta \|_\infty .1^{n-1}.\gamma$$

$$+$$

$$\vdots$$

$$+ g_n \cdot T_{\underline{i}}(z) \| d\theta \|_\infty .1^{n-1}.\gamma^n\Big]$$

for some $z \in X$, since X is compact

$$\leq 2A_1^2 \frac{\gamma}{1-\gamma} \cdot \sum_{|\underline{i}|=n} g_n \cdot T_{\underline{i}}(z)$$

where $A_1 = \max\{\| dg \|_\infty, \|1/g\|_\infty, \|d\theta\|_\infty\}$

$$\begin{aligned} &= 2A_1^2 \frac{\gamma}{1-\gamma} \bar{L}^n 1 \quad \text{by } (4{\cdot}2) \\ &= 2A_1^2 \frac{\gamma}{1-\gamma} \qquad \text{by Theorem 1·15 (i)} \\ &=: k_1. \end{aligned}$$

Note that the number of terms in (4·10) increases (polynomially) with n, yet their sum is bounded above by a number which is independent of n.

Successively differentiating (4·9) using the product and chain rules we see that for any $j - t \geq 1$ we have

$$\sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \leq 2(A_{j-t})^{j-t+1} \left(\frac{1}{1-\gamma}\right)^{j-t} =: k_{j-t} \ ,$$

where

$$A_{j-t} = \max\left\{\|d^l g\|_\infty, \|1/g\|_\infty, \|d^l\theta\|_\infty, \|d^{l+1}T_i\|_\infty : \ 1 \leq l \leq j-t, \ i \in I\right\}.$$

This completes the proof of Lemma B. $\square$

We continue the proof of Lemma 1·22.

For any $0 \leq j \leq k$ we have

$$
\begin{aligned}
&\| d^j(L_\theta^n w) \|_\infty \\
&\le \sum_{|\underline{i}|=n} \sum_{t=0}^{j} \binom{j}{t} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \| d^t(w \cdot T_{\underline{i}}) \|_\infty \quad \text{by (4·3)} \\
&= \sum_{t=0}^{j} \binom{j}{t} \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \| d^t(w \cdot T_{\underline{i}}) \|_\infty \\
&= \sum_{t=1}^{j} \Big[\binom{j}{t} \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \| d^t(w \cdot T_{\underline{i}}) \|_\infty \Big] + \sum_{|\underline{i}|=n} \| d^j(\psi_n \cdot T_{\underline{i}}) \|_\infty \| w \|_\infty \\
&\le \sum_{t=1}^{j} \Big[\binom{j}{t} \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \| d^t(w \cdot T_{\underline{i}}) \|_\infty \Big] + k_j \|w\|_\infty \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by Lemma B} \\
&\le \sum_{t=1}^{j} \Big[\binom{j}{t} \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \, (E_t \sum_{l=1}^{t} \| d^l w \|_\infty \, \gamma_0^{nl}) \Big] + k_j \|w\|_\infty \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by Lemma A} \\
&= \sum_{t=1}^{j} \Big[\binom{j}{t} (E_t \sum_{l=1}^{t} \| d^l w \|_\infty \, \gamma_0^{nl}) \sum_{|\underline{i}|=n} \| d^{j-t}(\psi_n \cdot T_{\underline{i}}) \|_\infty \Big] + k_j \|w\|_\infty \\
&\le \sum_{t=1}^{j} \Big[\binom{j}{t} (E_t \sum_{l=1}^{t} \| d^l w \|_\infty \, \gamma_0^{nl}) k_{j-t} \Big] + k_j \|w\|_\infty \quad \text{by Lemma B} \\
&\le F_j \sum_{t=0}^{j} \| d^t w \|_\infty \, \gamma_0^{tn} \qquad \text{where } F_j = \max\{k_j, \sum_{t=1}^{j} \binom{j}{t} E_t k_{j-t}\} > 0.
\end{aligned}
$$

So we have

$$
\begin{aligned}
\|L_\theta^n w\|_k &= \sum_{j=0}^{k} \|d^j(L_\theta^n w)\|_\infty \\
&\le \sum_{j=0}^{k} F_j \sum_{t=0}^{j} \|d^t w\|_\infty \gamma_0^{tn} \\
&\le C_1 \sum_{j=0}^{k} \|d^j w\|_\infty \gamma_0^{jn} \qquad \text{where } C_1 = \sum_{j=0}^{k} F_j > 0.
\end{aligned}
$$

This completes the proof of Lemma 1·22. □

Appendix B

In this appendix we give an adapted version of the original proof of Theorem 1·33. This result was originally proved, in the context of C^∞ Anosov diffeomorphisms, by de la Llave, Marco & Moriyón [34]. Our method has some similarities with the approach of Katok & Hasselblatt [25], page 610.

Theorem 1·33. *Suppose $T : X \to X$ is a piecewise $C^{\mathbf{r}}$, $1 < \mathbf{r} < \infty$, expanding Markov map of a partitioned manifold.*

Suppose $\Phi \in C^{\mathbf{k}}(X, \mathbf{R})$ for some $0 < \mathbf{k} = (k, \epsilon) \leq \mathbf{r} - 1$.

Suppose $W \in L^\infty(X, \mathbf{R})$ satisfies the real-valued cocycle equation

$$W(Tx) - W(x) = \Phi(x) \quad \text{a.e. } (m). \tag{4·11}$$

Then there exists $W' \in C^{\mathbf{k}}(X, \mathbf{R})$ such that

(i) $W' = W$ *a.e.* (m),

(ii) $W'(Tx) - W'(x) = \Phi(x)$ *everywhere.*

Proof. Since $W \in L^\infty(X, \mathbf{R})$, then by Theorem 1·32 there exists $W' \in C^{(0,\epsilon)}(X, \mathbf{R})$ such that $W = W'$ almost everywhere. So without loss of generality we will assume that $W \in C^{(0,\epsilon)}(X, \mathbf{R})$, and that (4·11) holds for all $x \in X$. We will first show that W is piecewise C^1, then that it is piecewise C^k, and finally that it is piecewise $C^{\mathbf{k}}$.

Let $\{T_i\}_{i \in I}$ be the family of inverse branches of T. These branches are all contractions, with contraction constant $\gamma \in (0,1)$ (see the remark after Definition 1·2).

Let $x \in X$. Choose $y \in X$ sufficiently close to x (close enough to ensure we may apply the same inverse branch, T_{i_1} say, to both points).

To the piece X_{i_1} (the domain of definition of T_{i_1}) we associate a sequence $(i_1, i_2, i_3, \ldots)$ such that for all $n \geq 1$ the composition $T_{i_n} \circ \ldots \circ T_{i_2} \circ T_{i_1}$ is well defined (on X_{i_1}). Since this sequence is fixed (for each piece X_{i_1}), it will cause no confusion to define $T^{-n} = T_{i_n} \circ \ldots \circ T_{i_2} \circ T_{i_1}$ (note that in Appendix A, where we sum over *all* such inverses, we use the notation $T_{\underline{i}}$ for such n-fold compositions).

For any $m \geq 1$ we have

$$W(x) - W(T^{-m}x) = \sum_{n=1}^{m} \Phi(T^{-n}x)$$

and

$$W(y) - W(T^{-m}y) = \sum_{n=1}^{m} \Phi(T^{-n}y).$$

Subtracting one equation from the other gives us

$$W(y) - W(x) = \sum_{n=1}^{m} [\Phi(T^{-n}y) - \Phi(T^{-n}x)] + W(T^{-m}y) - W(T^{-m}x). \tag{4·12}$$

Letting $m \to \infty$, we note that the distance $\rho(T^{-m}y, T^{-m}x) \to 0$, since the inverse branches are contractions. Hence $W(T^{-m}y) - W(T^{-m}x) \to 0$, since W is piecewise (Hölder) continuous. So (4·12) becomes

$$W(y) - W(x) = \sum_{n=1}^{\infty} [\Phi(T^{-n}y) - \Phi(T^{-n}x)]. \tag{4·13}$$

For any vector $v \in \tau_x X$, we will formally differentiate (4·13) with respect to the variable $y = x + sv$. Letting $s \to 0$ will give a formal series expression for the directional derivative of W at the point x in the direction v. We use $DW_v(x)$ to denote this directional derivative (even before showing that the series converges and that it actually exists). By hypothesis we know that the (total) derivatives $D_x T^{-n}$ and $D_x\Phi$ exist for all $x \in X$, and are continuous functions of x. Therefore the directional derivatives $DT_v^{-n}(x)$, $D\Phi_v(x)$ (at the point x and in the direction v) are equal to $D_x T^{-n}(v)$, $D_x\Phi(v)$ respectively.

Computing by the chain rule we have

$$\begin{aligned} DW_v(x) &= \sum_{n=1}^{\infty} D\Phi_{D_x T^{-n}(v)}(T^{-n}x).DT_v^{-n}(x) \\ &= \sum_{n=1}^{\infty} D_{T^{-n}x}\Phi(D_x T^{-n}(v)). \end{aligned}$$

Thus

$$\begin{aligned} \|DW_v\|_\infty &\leq \sum_{n=1}^{\infty} \|D\Phi\|_\infty \|DT^{-n}\|_\infty |v|_d \\ &\leq \|D\Phi\|_\infty |v|_d \sum_{n=1}^{\infty} \gamma^n \\ &< \infty. \end{aligned}$$

So the formal series expression for DW_v converges uniformly to a continuous function of x. But v was arbitrary, so all the directional derivatives DW_v are continuous. Therefore the total derivative DW exists and is continuous. That is, $W \in C^1(X, \mathbf{R})$.

So we have

$$\begin{aligned} DW &= \sum_{n=1}^{\infty} (D\Phi) \circ T^{-n}.DT^{-n} \\ &= \sum_{n=1}^{\infty} D(\Phi \circ T^{-n}). \end{aligned} \tag{4·14}$$

We will now show that W is piecewise C^k. The proof is by induction.

Our inductive hypothesis will be that $W \in C^j(X, \mathbf{R})$ for some $1 \leq j \leq k-1$, and satisfies the formula

$$D^j W = \sum_{n=1}^{\infty} D^j(\Phi \circ T^{-n}). \tag{4·15}$$

We note that the case $j = 1$ is true, by (4·14).

As before, we will obtain a formal expression for the directional derivative of $D^j W$ at the point x in the direction v, and then show this expression converges to a continuous function of x.

For arbitrary $x \in X$, $v \in \tau_x X$, we differentiate (4·15) to obtain

$$\begin{aligned} D(D^j W)_v(x) &= \sum_{n=1}^{\infty} D(D^j(\Phi \circ T^{-n}))_v(x) \\ &= \sum_{n=1}^{\infty} D_x^{j+1}(\Phi \circ T^{-n})(v). \end{aligned}$$

Recall from Lemma A of Appendix A (where we used the notation $T_{\underline{i}}$ instead of T^{-n}) that for any $\gamma_0 \in (\gamma, 1)$ and any $1 \leq j+1 \leq k$, there exists $E_{j+1} > 0$ such that, for all $n \geq 1$,

$$\begin{aligned} \|D^{j+1}(\Phi \circ T^{-n})\|_\infty &\leq E_{j+1} \sum_{l=1}^{j+1} \|D^l \Phi\|_\infty \gamma_0^{nl} \\ &\leq F_{j+1} \gamma_0^n \quad \text{for some } F_{j+1} > 0. \end{aligned} \tag{4·16}$$

Thus

$$\begin{aligned}\|D(D^j W)_v\|_\infty &\le \sum_{n=1}^{\infty} \|D^{j+1}(\Phi \circ T^{-n})\|_\infty |v|_d \\ &\le |v|_d F_{j+1} \sum_{n=1}^{\infty} \gamma_0^n \quad \text{by (4·16)} \\ &< \infty.\end{aligned}$$

So the infinite series expression converges uniformly to a continuous function. But the vector v was arbitrary, so all the directional derivatives $D(D^j W)_v$ are continuous. Therefore $D(D^j W)$ is continuous. That is, $W \in C^{j+1}(X, \mathbf{R})$. Moreover, (4·15) holds with j replaced by $j+1$. This completes the proof by induction.

So we have now established that $W \in C^k(X, \mathbf{R})$, and that

$$D^k W = \sum_{n=1}^{\infty} D^k(\Phi \circ T^{-n}). \tag{4·17}$$

It remains to show that $W \in C^{\prime k}(X, \mathbf{R}) = C^{(k,\epsilon)}(X, \mathbf{R})$.

Analogously to (4·16) there is a constant $G_k > 0$ such that for all $n \ge 1$ we have

$$\|D^k(\Phi \circ T^{-n})\|_\epsilon < G_k \gamma_0^n.$$

Consequently

$$\begin{aligned}\|D^k W\|_\epsilon &\le \sum_{n=1}^{\infty} \|D^k(\Phi \circ T^{-n})\|_\epsilon \quad \text{by (4·17)} \\ &\le G_k \sum_{n=1}^{\infty} \gamma_0^n \\ &< \infty.\end{aligned}$$

This completes the proof. $\square$

Appendix C

In this appendix we list the symbolic codes of all the extremal points of Ω_{19} with non-negative imaginary parts (see §3·4 for the definition of Ω_{19}). There are 61 such extremal points, which we list correct to 12 decimal places. By symmetry we can also recover those extremal points with negative imaginary part. In total Ω_{19} has 120 extremal points.

Note that all of the codes are Sturmian. That is, they are symbolic codes of ordered orbits. We believe that all of the extremal points of Ω_{19} are also extremal points of Ω.

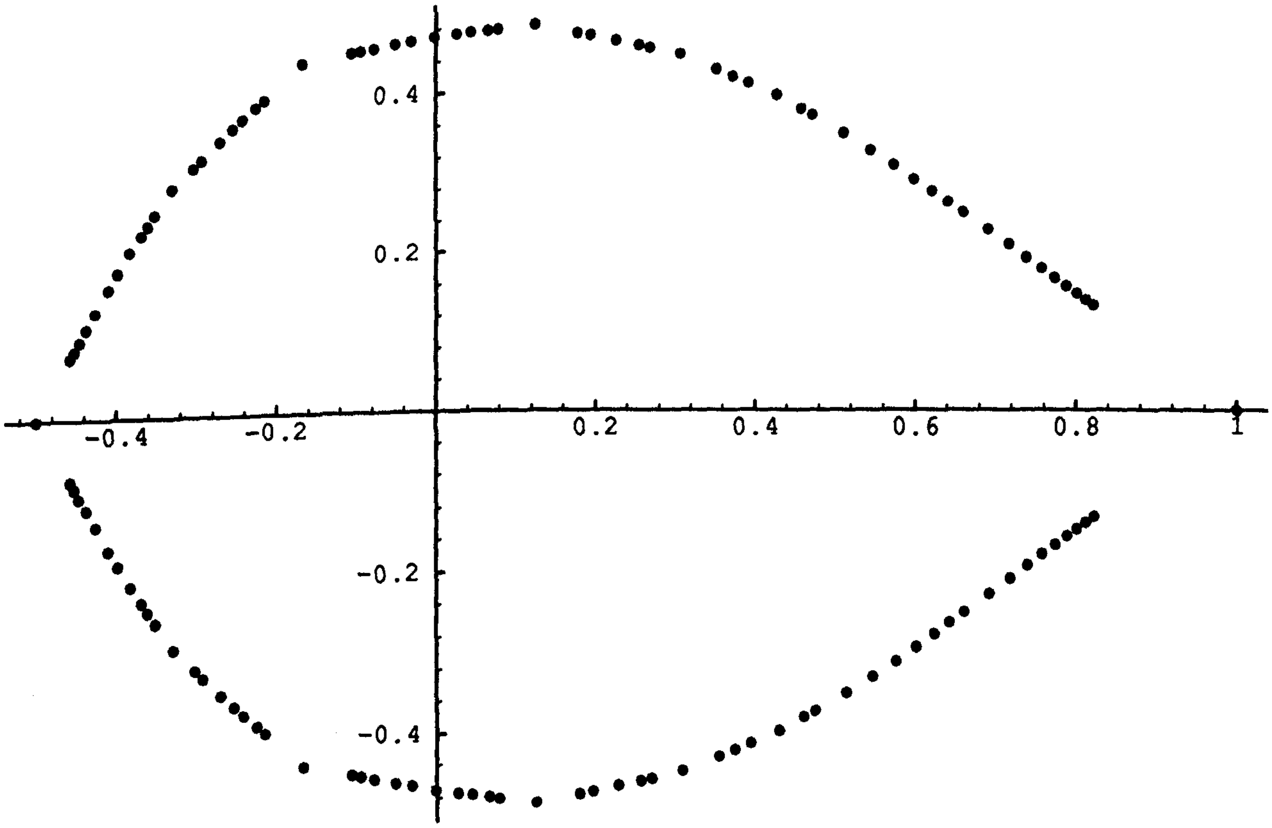
Period	Symbolic code	Barycentre
1	0	1
2	01	-0.5
3	001	$-0.166666666667 + 0.440958551844\,i$
4	0001	$0.125 + 0.484122918276\,i$
5	00001	$0.308387235944 + 0.443599069523\,i$
5	00101	$-0.329707054436 + 0.287689643293\,i$
6	000001	$0.428538113369 + 0.391976488612\,i$
7	0000001	$0.512531075989 + 0.345291450644\,i$
7	0001001	$-0.00140362486044 + 0.470885172778\,i$
7	0010101	$-0.3833036546409 + 0.20876673298\,i$
8	00000001	$0.574549340913 + 0.306148945021\,i$
8	00100101	$0.270194107446 + 0.346679649128\,i$
9	000000001	$0.622315350037 + 0.273906375018\,i$
9	000001001	$0.226902056657 + 0.462586823831\,i$
9	001010101	$-0.410179138253 + 0.162976012981\,i$

Period	Symbolic code	Barycentre
10	0000000001	$0.660308565458 + 0.247311280627\,i$
10	0001001001	$-0.051105078148 + 0.462336569705\,i$
11	00000000001	$0.691292345785 + 0.225189422959\,i$
11	00000100001	$0.373953453824 + 0.415632262833\,i$
11	00010001001	$0.044515779185 + 0.475883249288\,i$
11	00100100101	$-0.242101631147 + 0.372526749335\,i$
11	00101010101	$-0.426702170872 + 0.133465421068\,i$
12	000000000001	$0.717065205918 + 0.206588946532\,i$
12	001010010101	$-0.361038335018 + 0.24169754522\,i$
13	0000000000001	$0.738851246317 + 0.190773752459\,i$
13	0000001000001	$0.473773739402 + 0.36687768908\,i$
13	0000100010001	$0.195548258639 + 0.469254263091\,i$
13	0001001001001	$-0.077784910864 + 0.457444119809\,i$
13	0010010100101	$-0.29311131657 + 0.3240164004\,i$
13	0010101010101	$-0.438019320164 + 0.112957873233\,i$
14	00000000000001	$0.757514869067 + 0.177182456\,i$
14	00001000010001	$0.256004366153 + 0.45582389474\,i$
14	00100100100101	$-0.225950903032 + 0.387203897766\,i$
15	000000000000001	$0.773685278276 + 0.165386810788\,i$
15	000000010000001	$0.54560960847 + 0.324423836148\,i$
15	000100010001001	$0.065976174944 + 0.478088880836\,i$
15	001010101010101	$-0.446292196242 + 0.0979023639925\,i$
16	0000000000000001	$0.787832167746 + 0.155057877676\,i$
16	0000010000100001	$0.353464627893 + 0.424375928729\,i$
16	0001001001001001	$-0.0944514357351 + 0.454357229285\,i$
16	0010101001010101	$-0.398424211324 + 0.183011453934\,i$

Period	Symbolic code	Barycentre
17	00000000000000001	$0.800313672964 + 0.145940469508\, i$
17	00000000100000001	$0.599837723783 + 0.289081185291\, i$
17	00000100000100001	$0.393218917432 + 0.407285036394\, i$
17	00001000100010001	$0.178948700366 + 0.472754723444\, i$
17	00010010001001001	$-0.0306402627052 + 0.465858310809\, i$
17	00100100100100101	$-0.215490409728 + 0.396691370601\, i$
17	00101001010010101	$-0.351824709567 + 0.255225646268\, i$
17	00101010101010101	$-0.45261269921 + 0.086385660726\, i$
18	000000000000000001	$0.811407851056 + 0.137834386146\, i$
18	000100010010001001	$0.0266580284165 + 0.473940377714\, i$
18	001001010010100101	$-0.303277420255 + 0.313926201241\, i$
19	0000000000000000001	$0.821333987285 + 0.130580759738\, i$
19	0000000001000000001	$0.642311893071 + 0.259909366024\, i$
19	0000001000001000001	$0.459488896472 + 0.374804042626\, i$
19	0000100001000010001	$0.269789342506 + 0.452607258604\, i$
19	0001000100010001001	$0.078402141653 + 0.479359618122\, i$
19	0001001001001001001	$-0.105853966476 + 0.452242085786\, i$
19	0010010010100100101	$-0.253930324069 + 0.361644025924\, i$
19	0010100101010010101	$-0.369241665962 + 0.229565361695\, i$
19	0110101010101010101	$-0.457601269587 + 0.0772927065881\, i$

Appendix D

Figure 1. The 120 extremal points of Ω_{19}



References

[1] L. Ahlfors, *Möbius Transformations in Several Dimensions*, Ordway Professorship Lectures in Mathematics, University of Minnesota, 1981.

[2] C. Arteaga, Absolutely continuous conjugacy of expanding endomorphisms, *Ergod. Th. & Dyn. Sys.*, **13** (1993), 225–230.

[3] C. Arteaga, Differentiable conjugacy for expanding maps on the circle, *Ergod. Th. & Dyn. Sys.*, **14** (1994), 1–7.

[4] G. Atkinson, Recurrence of cocycles and random walks, *J. London Math. Soc.*, **13** (1976), 486–488.

[5] R. Berger, The undecidability of the domino problem, *Mem. Amer. Math. Soc. No.* 66, (1966).

[6] A. M. Blokh, Functional rotation numbers for one-dimensional maps, *Trans. Amer. Math. Soc.*, **347** (1995), 499–513.

[7] T. Bohr and D. Rand, The entropy function for characteristic exponents, *Physica*, **25D** (1986), 387–398.

[8] R. Bowen and C. Series, Markov maps associated to Fuchsian groups, *I.H.E.S. Publ. Math.*, **50** (1979), 401–418.

[9] P. L. Boyland, Bifurcations of circle maps: Arnol'd tongues, bistability and rotation intervals, *Comm. Math. Phys.*, **106** (1986), 353–381.

[10] S. Bullett and P. Sentenac, Ordered orbits of the shift, square roots, and the devil's staircase, *Math. Proc. Camb. Phil. Soc.*, **115** (1994), 451–481.

[11] R. M. Burton and J. E. Steif, Nonuniqueness of measures of maximal entropy for subshifts of finite type, *Ergod. Th. & Dyn. Sys.*, **14** (1994), 213–235.

[12] Z. N. Coelho-Filho, Entropy and ergodicity of skew-products over subshifts of finite type and central limit asymptotics, *Ph.D. Thesis*, Warwick University, (1990).

[13] D. Cohn, *Measure Theory*, Birkhäuser, Boston, 1980.

[**14**] M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic Theory on Compact Spaces*, Lecture Notes In Mathematics, **527**, Springer-Verlag, 1976.

[**15**] N. Dunford and J. T. Schwartz, *Linear Operators Part I*, Interscience, New York, 1958.

[**16**] M. J. Field, *Differential Calculus and its Applications*, Van Nostrand Reinhold, 1976.

[**17**] L. R. Ford, *Automorphic Functions* (2nd edition), McGraw-Hill, New York, 1951.

[**18**] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem of Diophantine approximation, *Math. Syst. Th.*, **1** (1967), 1–49.

[**19**] J. M. Gambaudo, O. Lanford and C. Tresser, Dynamique symbolique des rotations, *C. R. Acad. Sci. Paris*, **1299** (1984), 823–825.

[**20**] W. Geller and J. Propp, The fundamental group of a $\mathbb{Z}^2$ shift, *Ergod. Th. & Dyn. Sys.*, **15** (1995), 1091–1118.

[**21**] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers* (5th edition), Oxford, 1979.

[**22**] M. R. Herman, Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations, *I.H.E.S. Publ. Math.*, **49** (1979), 5–234.

[**23**] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis I*, Springer-Verlag, Berlin, Göttingen, Heidelberg, 1963.

[**24**] T. Kato, *Perturbation Theory*, Springer-Verlag, Berlin, 1976.

[**25**] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, 1995.

[**26**] A. Katok and R. J. Spatzier, Differential rigidity of hyperbolic abelian actions, *MSRI preprint*, (1992).

[**27**] A. Katok and K. Schmidt, The cohomology of expansive $\mathbb{Z}^d$-actions by automorphisms of compact, abelian groups, *Pacific J. Math.*, **170** (1995), 105–142.

[**28**] B. Kitchens and K. Schmidt, Periodic points, decidability, and Markov subgroups, *Lecture Notes In Mathematics*, **1342**, pp. 440–454, Springer-Verlag, 1988.

[29] O. E. Lanford, Entropy and equilibrium states in classical mechanics, *Lecture Notes In Physics*, **20**, pp. 1–113, (ed. A. Lenard), Springer-Verlag, 1973.

[30] F. Ledrappier, Un champ Markovien peut être d'entropie nulle et mélangeant, *C. R. Acad. Sci. Paris*, Ser. A **287** (1978), 561–562.

[31] A. Livsic, Homology properties of Y-systems, *Math. Zametki*, **10** (1971), 758–763.

[32] A. Livsic, Cohomology of dynamical systems, *Math. USSR Izvestija*, **6** (1972), 1278–1301.

[33] R. de la Llave, Invariants for smooth conjugacy of hyperbolic dynamical systems II, *Comm. Math. Phys.*, **109** (1987), 369–378.

[34] R. de la Llave, J. M. Marco and R. Moriyón, Canonical perturbation theory of Anosov systems and regularity results for the Livsic cohomology equation, *Ann. of Math.*, **123** (1986), 537–611.

[35] R. de la Llave and R. Moriyón, Invariants for smooth conjugacy of hyperbolic dynamical systems IV, *Comm. Math. Phys.*, **116** (1988), 185–192.

[36] R. Mañé, *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag, Berlin, Heidelberg, New York, 1987.

[37] J. M. Marco and R. Moriyón, Invariants for smooth conjugacy of hyperbolic dynamical systems I, *Comm. Math. Phys.*, **109** (1987), 681–689.

[38] J. M. Marco and R. Moriyón, Invariants for smooth conjugacy of hyperbolic dynamical systems III, *Comm. Math. Phys.*, **112** (1987), 317–333.

[39] B. Maskit, *Kleinian Groups*, Springer-Verlag, Berlin, 1988.

[40] F. Morgan, *Geometric Measure Theory*, Academic Press, San Diego, 1988.

[41] M. Morse and G. A. Hedlund, Symbolic Dynamics II. Sturmian Trajectories, *Am. J. Math.*, **62** (1940), 1–42.

[42] G. Mostow, Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms, *I.H.E.S. Publ. Math.*, **34** (1968), 53–104.

[43] S. Newhouse, J. Palis and F. Takens, Bifurcations and stability of families of diffeomorphisms, *I.H.E.S. Publ. Math.*, **57** (1983), 5–72.

[44] R. Nussbaum, The radius of the essential spectrum, *Duke Math. J.*, **37** (1970), 473–478.

[45] J. C. Oxtoby, Ergodic sets, *Bull. Amer. Math. Soc.*, **58** (1952), 116–136.

[46] W. Parry, Intrinsic Markov Chains, *Trans. Amer. Math. Soc.*, **112** (1964), 55–65.

[47] W. Parry, Instances of cohomological triviality and rigidity, *Ergod. Th. & Dyn. Sys.*, **15** (1995), 685–696.

[48] W. Parry, Skew-products of shifts with a compact Lie group, *Warwick Preprint*, (1995).

[49] W. Parry and M. Pollicott, The Livsic cocycle equation for compact Lie group extensions of hyperbolic systems, *Proc. London Math. Soc.*, to appear.

[50] W. Parry and M. Pollicott, *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics*, Astérisque 187–188, 1990.

[51] D. S. Passman, Group rings of polycyclic groups, *Group Theory: essays for Philip Hall*, pp. 207–256, (ed. K. W. Gruenberg and J. E. Roseblade), Academic Press, 1984.

[52] M. Pollicott, Meromorphic extensions of generalised zeta functions, *Invent. Math.*, **85** (1986), 147–164.

[53] M. Pollicott, The differential zeta function for Axiom A attractors, *Ann. of Math. (2)*, **131** (1990), 331–354.

[54] M. Pollicott, On the Ruelle-Tangerman theorem for zeta functions, *Proceedings of the European Conference on Iteration Theory*, Lisbon, 1991, 201–209.

[55] M. Reed and B. Simon, *Methods of Modern Mathematical Physics Vol. 1 : Functional Analysis, revised and enlarged edition*, Academic Press, 1980

[56] D. J. S. Robinson, *Finiteness Conditions and Generalized Soluble Groups, Part 2*, Springer-Verlag, Berlin, Heidelberg, New York, 1972.

[57] D. J. S. Robinson, *A Course in the Theory of Groups*, Springer-Verlag, New York, Heidelberg, Berlin, 1982.

[58] R. Robinson, Undecidability and nonperiodicity for tilings of the plane, *Invent. Math.* **12** (1971), 177–209.

[59] A. Rocha, Symbolic dynamics for Kleinian groups, *Ph.D. Thesis*, Warwick University (1994).

[60] H. L. Royden, *Real Analysis* (3rd Edition), Macmillan, 1988.

[61] D. Ruelle, The thermodynamic formalism for expanding maps, *Comm. Math. Phys.*, **125** (1989), 239–262.

[62] D. Ruelle, An extension of the theory of Fredholm determinants, *I.H.E.S. Publ. Math.*, **72** (1990), 175–193.

[63] K. Schmidt, The cohomology of higher-dimensional shifts of finite type, *Pacific J. Math.*, **170** (1995), 237–269.

[64] K. Schmidt, Cohomological rigidity of algebraic $\mathbb{Z}^d$ actions, *Ergod. Th. & Dyn. Sys.*, **15** (1995), 759–805.

[65] R. Sharp, Periodic points and rotation vectors for torus diffeomorphisms, *Topology*, **34** (1995), 351–357.

[66] M. Shub and D. Sullivan, Expanding endomorphisms of the circle revisited, *Ergod. Th. & Dyn. Sys.*, **5** (1985), 285–289.

[67] M. Spivak, *Calculus*, World Student Series Edition, Addison-Wesley, 1967.

[68] F. Tangerman, Meromorphic continuation of Ruelle zeta functions, *Ph.D. Thesis*, Boston University (1986).

[69] W. Thurston, *Geometry and Topology of 3-Manifolds*, section 5.9, mimeographed notes, Princeton University, 1978.

[70] P. Veerman, Symbolic dynamics and rotation numbers, *Physica*, **134A** (1986), 543–576.

[71] P. Veerman, Symbolic dynamics of order-preserving orbits, *Physica*, **29D** (1987), 191–201.

[72] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, Berlin, Heidelberg, New York, 1982.

[73] H. Wang, Proving theorems by pattern recognition II, *Bell System Tech. J.* **40** (1961), 1–41.

[**74**] B. A. F. Wehrfritz, *Infinite Linear Groups*, Springer-Verlag, Berlin, Heidelberg, New York, 1973.

[**75**] K. Ziemian, Rotation sets for subshifts of finite type, *Fund. Math.*, **146** (1995), 189–201.